MACRO
CONSULTING, INC.
Marketing Research / Advanced Analytics

McCULLOUGH LAW'S:
First Principals of Commercial Data Analysis

In a career that is quickly, too quickly, approaching 30 years in length, I have stumbled upon a series of general principles that have been proven, usually by my not employing them, to successfully guide the earnest analyst as he or she tentatively picks his or her way through that dark and tangled forest that we often refer to as a commercial data set; all this in his or her quest for the holy grail of data analysis, Truth.

Author: Paul Richard "Dick" McCullough

McCULLOUGH LAW'S:

First Principals of Commercial Data Analysis

In a career that is quickly, too quickly, approaching 30 years in length, I have stumbled upon a series of general principles that have been proven, usually by my not employing them, to successfully guide the earnest analyst as he or she tentatively picks his or her way through that dark and tangled forest that we often refer to as a commercial data set; all this in his or her quest for the holy grail of data analysis, Truth.

Thus, this article humbly serves to summarize these Laws and their accompanying theorems and corollaries in much the same way as Maxwell summarized the laws of electricity and magnetism over a hundred and forty years ago. Yes, I know. I'm a bit behind.

McCullough's First Law of Statistical Analysis

If you still have trouble understanding the difference between accepting the null hypothesis and failing to reject it, you will find the First Law extremely useful. Forget all that conceptual nonsense and apply the First Law with vigor. You'll be fine.

> *If the statistics say an effect is real,*
> *it probably is.*

> *If the statistics say an effect is not real,*
> *it might be anyway.*

This is true because none of you bother to look at beta errors (don't worry, I don't either). I mean, who's got the sample size, anyway? If you do worry about such things as beta errors and power curves (and you know who you are), either you are an academic (and should have stopped reading this article long ago) or you are in desperate need of an appropriate 12 step program. When in doubt, see your nearest HR representative.

Douglas MacLachlan, a distinguished professor at the University of Washington, was kind enough to let me repeat a law he often shares with his graduate students, which captures the spirit and intent of my First Law very well:

MacLachlan's Law

> *Torture any data set long enough,*
> *and it will confess.*

The point being, of course, DON'T QUIT! Data don't yield themselves up to the dashing data monger like some chambermaid from a gothic novel. No, data are reluctant lovers that must be coaxed and wooed (and occasionally slapped around a bit). The successful data analyst is the one

who is tenacious. Remember, data are not people. If they say no (and even if they mean it), you don't have to listen. Pretend they meant yes. Barge ahead.

It turns out, rather unfortunately, that Professor MacLachlan means exactly the opposite to the above when he quotes MacLachlan's Law.   Professor MacLachlan's point is you can artificially manufacture from your data set virtually any story you want by exhaustive and indiscriminant analysis (not to be confused with discriminant analysis which is an entirely different cup of tea). But that quickly leads us into a philosophical discussion of theory-driven versus data-driven models. Nobody wants to go there, believe me. Let me just say if you follow, with the dedicated fervor of an English soccer hooligan, McCullough's Second Law (see below), or more specifically, the Corollary to the Second Law (see below below), you will safely steer clear of any trouble with MacLachlan's Law (as interpreted by MacLachlan).

McCullough's Second Law of Statistical Analysis

> *Never, ever confuse significance with importance.*

Imagine a battery of 100 brand imagery statements. Now imagine the master you serve wants you to test for significant differences between right handed respondents and left handed respondents (don't pretend you haven't faced similarly mind-numbing requests with a toothy smile and a perky "good idea, sir!"). Dutifully, you conduct 100 pairwise tests. Not surprisingly, you find 5 statements have significantly different mean ratings for right and left handed persons.   This finding is likely unimportant for two entirely different reasons.

Of course, 100 pairwise tests are apt to generate some statistically significant differences by accident. I mean, by definition the tests are only accurate 95% of the time, right? So they're inaccurate 5% of the time.   There are easy ways around this. Go look them up (see if you can find Fisher's pooled alpha-obscure but cool). In the meantime, ignore these 5 differences because they are very probably spurious.

These differences are likely to be unimportant in a second, more important way. At least in my example, these differences are likely to not tell you anything that you can use in your business practice. Suppose left handed respondents truly do think Brand A is very slightly more "frivolous" (or "playful" or "precocious" or "insouciant," ad nasuem) than Brand B (7.8 vs. 7.7 out of 10, say). Now what? Shall we hang a 10 million dollar ad campaign on this statistically significant finding? How?

The careful reader will note that I said these 5 differences are likely to be unimportant. What if all 5 differences are consistent with one another? That is, what if they provide face validity for each other? Tell a compelling story that makes sense and is actionable? In that unlikely event, skip the next corollary and law and proceed directly to McCullough's Third Law (do not pass Go; do not collect $200).

Corollary to the Second Law

> *If it doesn't make sense,don't do it.*

The Corollary to the Second Law could also be called the First and Most Important Principle of Confirmatory Analysis because it is the cornerstone of confirmatory modeling, such as structural equations or confirmatory factor analysis. Statistical principles, like religious ones, should stand the test of common sense. Otherwise, you run the risk of committing atrocious crimes against humanity (or numbers) that violate the very principles you sought to uphold. To be blunt, if the sign of your regression coefficient is opposite to what you know to be true, deal with it! Check for collinearity,

double check the data set for coding errors, toss the bugger out if you have to. But don't just leave the absurdity in your model because the data said so. The data didn't say so any more than a rock tells you to throw it through a window.

Don't get me wrong. It's alright to explore a data set. Try things out. Play around. Experiment. Just examine your conclusions with a healthy dose of reality, aka, cynicism. The first time you try telling a grizzled veteran of the trenches that his or her sales are not affected by a key competitor's drop in price, you'll appreciate the wisdom (painfully earned, I might add) of the Corollary to the Second Law.

The Rolling Stones Law

> *You can't always get what you want*
> *But if you try, you just might find*
> *You get what you need.*

Besides being lyrics from a really cool, or to be more current, blingin, song, the Rolling Stones Law reminds us that we are not, despite being the high priests and holy gatekeepers of the Information, in control (and given my track record, this is a very good thing). It is natural to approach a data set with some preconceived ideas about what is going on. But we can't let those prejudices influence our search through the data (remember MacLachlan's interpretation of MacLachlan's Law?). We must accept what the Data God gives us, even if it isn't the answer that our president's wife thinks that it should be. We must strive for Zen-like detachment and accept what is. Mick and the gang assure us it will be enough. Keep the faith, brothers and sisters!

McCullough's Third Law of Statistical Analysis

> *If you can't tell what it is,*
> *it ain't art.*

Apologies to Jackson Pollack, et al. All the hoity toity statistics are nice, particularly as entertainment for us quant jocks, but the ultimate goal, as stated earlier, is to find Truth. If you have a seemingly random collection of statistically significant differences, a la Pablo Picasso, please see the second law above because you haven't found Truth. If you have a coherent picture, a la Leonardo daVinci, confirmed and/or corroborated by numerous independent data points, you're likely to be barking up Truth's tree, whether you have official significance or not. Remember:

McCullough's Law of Small Samples

> *Give me a sample small enough,*
> *And all means will be statistically equal.*

Sample size is like dirt on a window pane. The smaller the sample size, the dirtier the window. If there are two sprawling, old oak trees majestically parked side by side in front of the window, a small sample size will make them harder to see but it won't have any affect on whether or not there are two trees instead of one. If you stare through five dirty windows and think you see the hazy outline of two big oak trees each time, there's probably two different trees out there. Through one window alone, it might just be random patterns in the dirt.

Not wishing to confuse you, but there is a flip side to the Law of Small Samples:

McCullough's Law of Large Samples

Give me a sample large enough,
And all means will be statistically significant.

The problem with data analysis (survey data, anyway) is that it comes with some assumptions that are usually well hidden. And, in the real world, these assumptions are almost never fully realized. As sample size gets larger, the window gets clearer (see above). And when staring at an apparently statistically significant difference, we know we're seeing something. The question is what.

Let's say we did an online survey with a million people. They rated 5 brands on preference. Just preference. There may be the slightest bit of fatigue that sets in by brand number 5. Or maybe some respondents have a very small computer screen and have to scroll over to see brand 5. Scrolling annoys them and they subconsciously take it out on brand 5. These very tiny influences may show up as statistically significantly lower brand preference ratings for brand 5 even if brand 5 is equally preferred to the other brands. The large sample size allows us to identify a non-random effect. It just might not be the effect we are expecting. Can you say "measurement error?"

It's sad, isn't it? All of us who enter the Brotherhood, do so with the fervent desire of first finding, then standing on firm ground. It's what attracts us in the first place. All these rules, equations, tedium. There must be a reward. And that reward is certainty. Alas, no.

But let's not dwell on this too long, ok?

McCullough's Law of Cluster Analysis

The name of the cluster
is always more important
than the cluster

Keep in mind that finding Truth (or your somewhat imperfect version of it) won't do much good if you can't explain it to someone else, most likely someone in marketing. You can't expect a marketing person to digest a series of distribution comparisons. If he or she could, they wouldn't be in marketing. No, the sum total of learning to be gleaned from your three weeks of segmentation analysis will be contained in the names of your clusters, whether you like it or not. So name them carefully. Ditto for any other golden nuggets of knowledge you wish to share with the unwashed masses. In sum: Be brief. Be simple. Be clear.

Research dollars come straight off the bottom line. If we don't help the guys on the firing line make more money, then we've helped them make less. So let's bend over backwards to help them "get it." Speak plain English (or whatever).

My last two laws may be the most important. They are both yellow caution flags telling the avid analyst to beware the twin pitfalls of enthusiasm and earnestness. I've saved them for the Big Finish (drum roll, please).

McCullough's Law of Statistical Fashion

Give a kid a hammer,
and the world becomes a nail.

Been there, done that. Like a kid at Christmas with a new toy, the analyst, armed with a recently mastered (or not) statistical technique, sees every marketing problem as an opportunity to practice his newly acquired vocation. We analysts must view the ever increasing portfolio of powerful statistical

techniques available to us in user-friendly drop down menus (and soon portable to your cell phone and/or wrist watch), as a selection of beautiful arrows in our analytical quiver. Each arrow serves its own unique purpose. The competent analyst will be prepared to use whatever arrow is right for his or her business problem, not vice versa.

McCullough's Law of Marketing Impact

Any direction is
Better than no direction.

Analysts seek Truth. It is what we do. Finding it with certainty is problematic, however, for numerous reasons. If marketing is a voyage, research is the compass. If you were lost in a mountain range full of iron ore, would you sit still until you were certain your compass was accurate, or would you start walking while there was still light? If you're a true stat geek, you'll be frozen solid by morning while the marketing guys will be sipping lattes at an outside café.

When you study a data set, the easiest (and most cowardly) path is "we didn't see anything significant." Sometimes, if you've done an extraordinarily poor job designing your study in the first place, this may be true. But most of the time, there are stories in there. Some of them take a little more digging to get to, but they are there. Dig. Paint a picture. Suffer for your art. The more independent data points you can find that sing the same song, the more likely you are at least getting warm (remember the dirty windows). Businesses rarely succeed by inaction. And most marketers rarely remain marketers through inaction. Get in there and help. So you're stretching the data set a little bit. It's not like you're pulling the wings off butterflies. It's ok. Data don't have feelings. Data don't cry. Search your data with an open mind, a pure heart and a ruthless spirit.

Do you really want those marketing guys stumbling down the mountain without a compass?

Additional Reading

Campbell, Donald T. and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Rand McNally College Publishing Company, Chicago, 1963.

Cohen, Jacob, *Things I Have Learned (So Far)*, American Psychologist, December 1990.

Huff, Darrell, *How to Lie with Statistics*, W. W. Norton & Company, New York, 1954.

McCloskey, Deirdre N., *The Vices of Economists-The Virtues of the Bourgeoisie; Chapter: The Irrelevance of Statistical Significance*, Amsterdam University Press and University of Michigan Press, 1997.

McCullough, Dick, *Three and a half Steps to Statistical Success*, Quirk's Marketing Research Review, December 1997.

A version of this paper was published in the January 2007 issue of QUIRK'S MARKETING RESEARCH REVIEW

MACRO
CONSULTING, INC.

We are an independent **marketing research consulting firm** dedicated to helping you make the most informed, insightful marketing decisions possible. We specialize in technology, consumer, and new product research, and are well recognized for our **State-of-the-Art Research** techniques**.**

Ultimately, we provide more than just technical expertise.
We focus on developing **pragmatic solutions** that will have a positive impact on the profitability of our clients.

CONTACT US:
Telephone: 650-823-3042
General Inquiries:
info@macroinc.com

Advanced Analysis Inquiries:
analysis@macroinc.com

richard@macroinc.com

www.macroinc.com