Averages, Schmaverages!
Why You Shouldn't Use Averages for Performance Prediction

A White Paper from Responsive Systems

by
Dr. Bernie Domanski

Abstract

Statistics have great power; they can be used to convey the true, and sometimes not obvious, meaning behind a particular environment or a set of events. But used incorrectly under the banner of 'science and mathematics', statistics can produce misleading and disastrous results.

Our objective in this paper is simple: namely we will explain why averages of averages, or moving averages, or just plain old averages, should not be used for any statistical inference or performance prediction. We'll explain what the different types of averages are, with examples from a variety of disciplines, including sports and the stock market, and we'll supply the mysterious math behind it all. We hope that the inadequacies of averages will be illustrated to your satisfaction, and hope that you will be armed to better understand what different performance packages are actually supplying.

Preconceived Notions of Streaks, Slumps and the Law of Averages[1]

There is nothing sportscasters like to talk about more than streaks, slumps and the law of averages. And many sportscasters show their ignorance of statistics when they do so. Should you ignore streaks, slumps, and the law of averages?

Streaks and slumps are related. A player who does a good thing a lot over some period of time is said to be on a streak (or is "hot"), while a player who does something poorly over a period of time is in a slump (or is "cold"). These concepts are very popular in the sports media. Unfortunately, they are of little meaning and of less use.

What does a streak tell you? In hockey, for instance, scoring streaks are often mentioned. Why is it important to know that Player X has five goals in his last four games? While it is true that a common predictor of future performance is past performance, any streak mentioned is of such *a short time period* as to be of no predictive use whatsoever. That is the main problem with streaks and slumps; you can select any period of x number of games to examine. Streaks and slumps are merely the result of choosing your endpoints carefully. If we map the analogy over to database performance, the same must be true … we must choose our endpoints carefully.

Let's compare two players, over a period of 10 games. Player A scores a goal every second game, for a total of five goals. Player B goes scoreless for the first three games, and then scores a single goal in each of the next four games, then goes scoreless for the final three games, for a total of four goals. Player A scored five goals compared to Player B's four, but you can bet that someone in the media will mention Player B's "four-game scoring streak" at some point. Attention is directed to where it should not be.

Really, the only appropriate time period to examine is the full season, and only for evaluation of what has happened. This is the basic unit of sports. A team's record over the entire season is used to determine if it makes the playoffs. Thus, we should use a player's full season when evaluating him. It is a fact that all players will go through a series of "streaks" and "slumps" during a season, on his way to his final level of production. Breaking down performances into arbitrary sets of games has no value.

Streaks and slumps are also very subjective in their definition; that is, there is no actual definition. The judgment on whether a player is on a streak or in a slump depends on the perceived quality of that player. The subjectivity of streaks and slumps is best demonstrated when one is "broken". For instance, say a player has gone 20 games without scoring a goal, and this is considered a slump for that player. The media will surely mention his "20-game scoreless drought" or some such thing. In

[1] http://www.hockeyzoneplus.com/puckerings/puck004.htm

his next game, the player scores a goal, and it will be mentioned that he has "broken out" of his slump. This is supposed to mean he is no longer in a slump. But if you look at it, he has still only scored one goal in his last 21 games, which would also be considered a slump. He has broken out of a slump, yet he is still in a slump. This is arbitrary and silly.

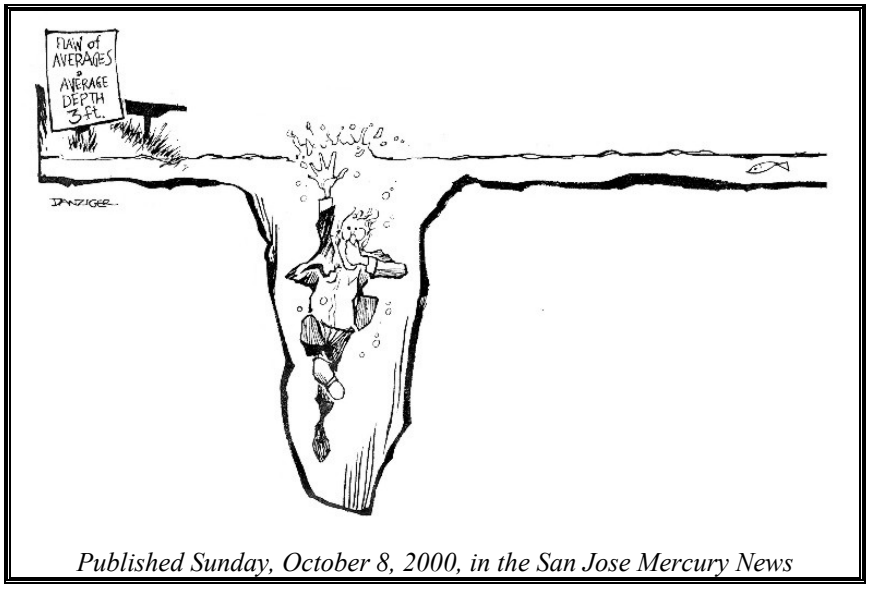
The Flaw of Averages[2]

> *"The only certainty is that nothing is certain."*
> *- Roman scholar Pliny the Elder*

Said some 2000 years ago, it's a safe bet Pliny the Elder would still be right. The Information Age, despite its promise, also delivers a dizzying array of technological, economic and political uncertainties. This often results in the "*Flaw of Averages*", a fallacy as fundamental as the belief that the earth is flat.

The Flaw of Averages states that: ***plans based on the assumption that average conditions will occur are usually wrong***.

Consider the statistician who drowned while fording a river that was, on average, only three feet deep.



Published Sunday, October 8, 2000, in the San Jose Mercury News

In real life, the *flaw* continually gums up investment management, production and performance planning and other seemingly well-laid plans. The Flaw of *Averages is one of the cornerstones of Murphy's Law (*What can go wrong does go wrong).

Computers can overcome this problem by bombarding our plans with a whole range of inputs instead of single average values. Today, this technique, known as *simulation*, is

[2] http://www.stanford.edu/~savage/flaw/Article.htm

at the center of such diverse activities as Wall Street investing and military defense planning, and of course, performance prediction.

But back to the flaw, and an area that's important to all of us: investing for the future.

Suppose you want your $200,000 retirement fund invested in the Standard & Poor's 500 index to last 20 years. How much can you withdraw per year? The return of the S&P has varied over the years but has averaged about 14 percent per year since its inception in 1952. You use an annuity workbook in your spreadsheet that requires an initial amount ($200,000) and a growth rate for the fund. "*I need a number*," you say to yourself, so you plug in 14 percent. Now you can play with the annual withdrawal amount until your money lasts exactly 20 years. If you do this you will be pleased to find that you can withdraw $32,000 per year. (see Figure A below).

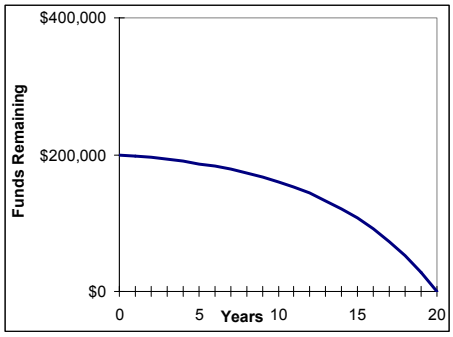


Figure A. Funds remaining with annual withdrawal of $32,000 , assuming 14% return every year

Even if the return fluctuates in the future, as long as it averages 14 percent per year, the fund should last 20 years, right? **Wrong**! Given typical levels of stock market volatility, there are only slim chances that the fund will survive the full time. The following charts simulate this retirement strategy with actual S&P 500 returns starting in various years.

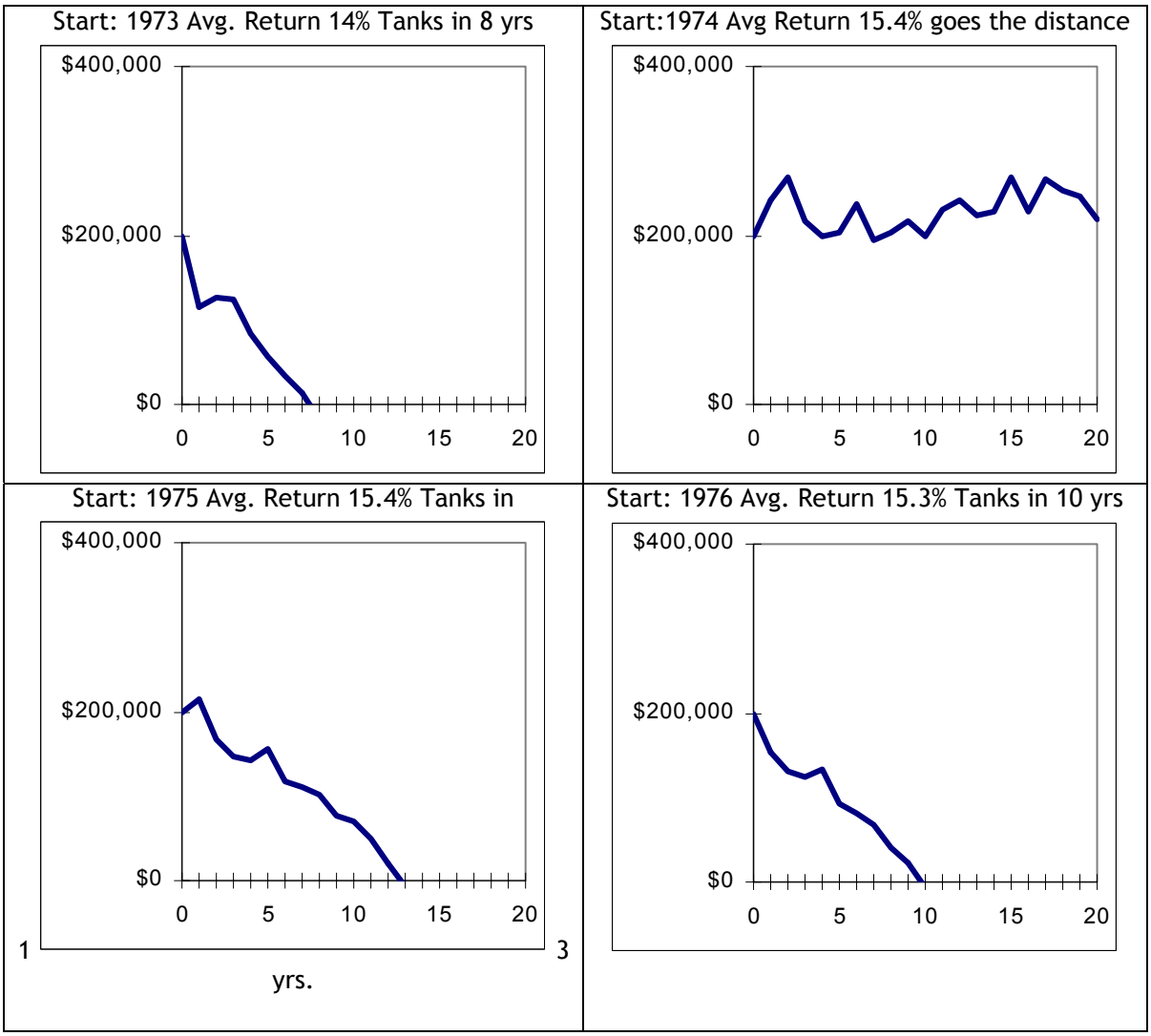


Figure B. Simulated Fund performance if started in various years.

Notice that the level of average returns over any particular 20-year period is no guarantee of success. The real key is to get off to a good start, which is what separates 1974 from its neighbors.

For this example, the Flaw of Averages states that if you assume each year's growth at least equals the average of 14 percent, there is no chance of running out of money. But if the growth fluctuates each year but averages 14 percent, you are likely to run out of money.

These results are not the result of a rigorous scientific study, and should not be used for making investment decisions, but they should at least have you asking yourself:

Why isn't someone doing something about this? People are. One of the first was William F. Sharpe, a Nobel laureate in Economics, who recently left Stanford to spend full time ***simulating*** retirement benefits. "*I expected people to question the specifics of our simulation algorithms*," reflects Sharpe about the launch of his Palo Alto-based Financial Engines Inc., "*but to my surprise, everyone else out there was just plugging in averages.*" (as in Figure A)

The Flaw of Averages distorts everyday decisions in many other areas. Consider the case of a Silicon Valley product manager who has just been asked by his boss to forecast demand for a new-generation microchip.

"*That's difficult for a new product*," responds the product manager," *but I'm confident annual demand will be between 50,000 and 150,000 units*."

"*Give me a number to take to my production people,*" barks the boss. "*I can't tell them to build a facility with a capacity of between 50,000 and 150,000 units*!"

So the product manager dutifully replies: "*If you need a single number, the average is 100,000.*"

The boss plugs the average demand and the cost of a 100k capacity facility into a spreadsheet. The bottom line is a healthy $10 million, and he reports this to his board as the average profit to expect. Assuming that demand is the only uncertainty, and that 100,000 is the correct average, then $10 million must be the best guess for profit. Right? **Wrong**! The Flaw of Averages ensures that *average profit* will be less than the profit associated with the *average demand*. Why? Lower-than-average demand clearly leads to profit of less than $10 million. That's the downside. But greater demand exceeds the capacity of the plant, leading to a maximum of $10 million. There is no upside to balance the downside!

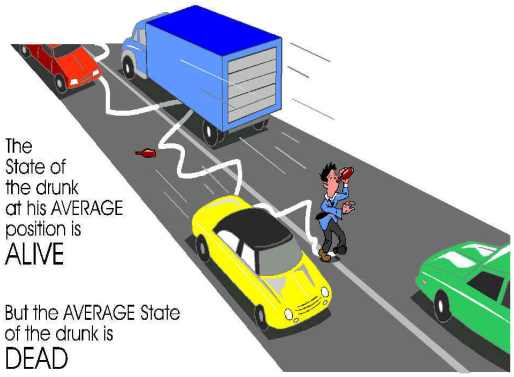
This leads to a problem of Dilbert-like proportion: the product manager's correct forecast of average demand leads to an incorrect forecast of average profit, so he gets blamed for giving the correct answer.

A computerized prescription for the Flaw of Averages is *Monte Carlo Simulation*, first used for modeling uncertainty during development of the atomic bomb. It generates thousands of scenarios covering all conceivable real world contingencies in proportion to their likelihood.

In the 1950s, Harry Markowitz, a brash young graduate student at the University of Chicago, dealt another blow to the flaw. "*I was reading the contemporary investment theory, which was strictly based on averages*," recalls Markowitz. "*I said to myself: `this can't be right.'*" His resulting portfolio theory, which was based on both risk and average outcomes, revolutionized Wall Street and won him a Nobel Prize. Markowitz also devoted much of his career to designing *simulation* systems.

Simulation-based acquisition is used routinely in the military. Its instigator was William J. Perry. In 1996, while at the Pentagon, Perry issued a directive stating that models and simulations must be used to reduce the time, resources and risks of the acquisition process. Perry says in retrospect: "*With tens of thousands of uncertainties, it was just a perfect application for simulation.*"

A dramatic example of the savings that resulted from Perry's directive is related by John D. Illgen of Santa Barbara-based *Illgen Simulation Technologies* Inc., who says: "*In response to improvements in foreign weapon systems, the Navy was preparing to spend tens of millions of dollars to upgrade its shipboard defensive systems. With a $250,000 simulation we were able to show that the present defensive system was adequate to meet the increased threat.*"



While many of today's managers still cling tenaciously to "*flat earth*" ideals, the innovators are abandoning averages and facing up to uncertainty. Those who dare discover a *New World* of managerial tools including simulation, decision trees, and other real options.

And what happens when one of these innovators is confronted by someone cloaking themselves behind a single number? The story of the emperor's new clothes says it all.

Moving Average[3]

Moving Averages are one of the most popular and easy to use tools available to the technical analyst. By using an average of stock prices, for example, moving averages *smooth* a data series and make it easier to spot trends. This can be especially helpful in volatile markets.

A moving average (MA) is an average of data for a certain number of time periods. It "*moves*" because for each calculation, we use the last *n* number of data points. There are two major types of Moving Averages: "Simple" and "Exponential".

Simple Moving Average
A simple moving average (SMA) is formed by finding the average price of a currency or commodity over a set number of periods. For example, the closing stock price is used to compute the moving average. For example: a 5-day moving average would be calculated by adding the closing prices for the last 5 days and dividing the total by 5.

[3] http://www.forex-training.com/moving_average.htm

$$10 + 11 + 12 + 13 + 14 = 60$$

$$60 \div 5 = 12$$

A moving average *moves* because as the newest period is added, the oldest period is dropped. If the next closing price in the average is 15, then this new period would be added and the oldest day, which is 10, would be dropped. The new 5-day moving average would be calculated as follows:

$$11 + 12 + 13 + 14 + 15 = 65$$

$$65 \div 5 = 13$$

Over the last 2 days, the moving average moved from 12 to 13. As new days are added, the old days will be subtracted and the moving average will continue to move over time. Moving averages are *lagging indicators* and will always be behind the current price. They fit in the category of *trend following*. When prices are trending, moving averages work well. However, when prices are not trending (e.g. when there is a peak or an exceptionally "low" stock price), moving averages do not work.

Exponential Moving Average

To reduce the lag in simple moving averages, analysts sometimes use exponential moving averages, or exponentially weighted moving averages. Exponential moving averages reduce the lag by applying more weight to *recent prices relative to older prices*. The weighting applied to the most recent price depends on the length of the moving average. The shorter the exponential moving average is, the more weight that will be applied to the most recent price. For example: a 10-period exponential moving average weighs the most recent price 18.18% and a 20-period exponential moving average weighs the most recent price 9.52%. The method for calculating the exponential moving average is fairly complicated. The important thing to remember is that the exponential moving average puts more weight on *recent* prices - it will react faster to recent price changes than a simple moving average.

Exponential Moving Average (EMA) Calculation

The formula for an exponential moving average is:

$$X = (K \times (C - P)) + P$$

where:
X = Current EMA
C = Current Price
P = Previous period's EMA*
(*A Simple Moving Average is used for first period's calculation)
K = Smoothing constant

The smoothing constant applies the appropriate weighting to the most recent price relative to the previous EMA. The formula for the smoothing constant is:

K = 2/ (1+N)
N = Number of periods for EMA

For a 10-period EMA, the smoothing constant would be .1818.

$$\frac{2}{(\text{Time periods} + 1)} = \frac{2}{(10 + 1)} = .1818 \ (18.18\%)$$

The EMA formula works by weighting the difference between the current period's price and the previous period's EMA and adding the result to the previous period's EMA. There are two possible outcomes: the weighted difference is either positive or negative.

If the current price (C) is higher than the previous period's EMA (P), the difference will be positive (C - P). The positive difference is weighted by multiplying it by the constant ((C - P) x K) and the answer is added to the previous period's EMA, resulting in a new EMA that is higher ((C - P) x K) + P.

If the current price is lower than the previous period's EMA, the difference will be negative (C - P). The negative difference is weighted by multiplying it by the constant ((C - P) x K) and the final result is added to the previous period's EMA, resulting in a new EMA that is lower ((C - P) x K) + P.

Great care must be taken when using exponential moving averages; the trick is to select an 'optimal' number of periods. Considerable research has been done in this area within the stock market, and in particular, *Moving Average Convergence / Divergence (MACD)* developed by Gerald Appel. MACD is a simple indicator that is available for long term trending. MACD is calculated by subtracting the 12-period EMA of a given metric from its 26-period EMA. A 9-period EMA of the *MACD itself* is usually plotted over this line as a *signal* or *trigger line*. By using moving averages, MACD has trend-following characteristics, but it still not an absolute predictor of performance.

Law of Averages[4]

[4] http://stat-www.berkeley.edu/~stark/Java/lln.htm

So while Moving Averages, and in particular, *exponential* moving averages are good indicators of past trends, it is usually the infamous *Law of Averages* that is often misused.

The Law of Averages[5] (From Wikipedia, the free web encyclopedia) is a lay term used to express the view that eventually, everything "*evens out*." For example: Two people who drive cars over a long period of time will have roughly the same number of accidents. The more children you have, the more likely you will have an equal division of boys and girls. The longer you flip a coin, the more likely the number of heads and tails will equalize.

In mathematical jargon, It says that *the average of independent observations of random variables that have the same probability distribution is increasingly likely to be close to the expected value of the random variables as the number of observations grows*. In English, this simply means that if all of the possible outcomes are equally likely, then the expected value is the average.

More precisely, if X1, X2, X3, . . . , are independent random variables with the same probability distribution, and their expected value, *E(X)*, is their common expected value, then for every number E > 0, P{|(X1 + X2 + . . . + Xn)/n - E(X) | < E} converges to 100% as n grows. This is equivalent to saying that the sequence of sample means X1, (X1+X2)/2, (X1+X2+X3)/3, . . .

In other words, the Law of Averages says that I'll eventually be right! [6]

While the general belief in the law of averages is why people gamble in Las Vegas on the belief that they will "sooner or later break even", the *law of large numbers* is why casinos around the world make billions of dollars. The law of large numbers states that a large sample of a particular probabilistic event will tend to reflect the underlying probabilities. For example, after tossing a "fair coin" 1000 times, we would expect the result to be approximately 500 heads results, because this would reflect the underlying .5 chance of a heads result for any given flip.

However, it is important that while the *average* will move closer to the underlying probability, in absolute terms *deviation from the expected value will increase*. For example, after 1000 coin flips, we might see 520 heads. After 10,000 flips, we might then see 5096 heads. The average has now moved closer to the underlying .5, from .52 to .5096. However, the *absolute deviation* from the expected number of heads has gone up from ***20*** *to* ***96***.

There are two common ways to misunderstand and misapply the law of large numbers:

[5] http://en.wikipedia.org/wiki/Law_of_averages
[6] http://www.yergler.net/averages/

- "*If I flip this coin 1000 times, I will get 500 heads results*." ***False.*** While we expect *approximately* 500 heads, it is not the case that we will always get **exactly** 500 heads results. Similarly, getting 520 heads results is not conclusive proof that the coin's true probability of getting heads on a single flip is .52

- "*I just got 5 tails in a row. My chances of getting heads must be very good now*." ***False.*** Many probabilistic events are independent of one another, which means the result of one event does not in any way influence the outcome of another. Coin flips are independent events. The coin does not "remember" what it has flipped previously and self-adjust to get an overall average result. The coin is not "due" for a heads. The probability remains .5 for each individual flip. A belief in this fallacy can be devastating for amateur gamblers. The thought that "*I have to win soon now, because I've been losing and it has to even out*" can encourage a gambler to continue to bet more than they can afford.

Database Performance Tools

Pool Advisor for DB2 (BMC) Uses a collection agent to return a snapshot at one minute intervals of global system-wide resource usage of recent system activity. BMC claims that similar products use a traditional tuning method based on short collections or "snapshots" in time that result in configuration decisions based only on conditions that exist at the time of the snapshot – in point of fact, this is exactly what the Pool Advisor is doing. If tuning is based on longer collection samples, the highs and lows of resource use can be average, and that can result in performance problems during peak usage – the reader is asked to refer back to the previous sections of this paper for evidence of averages misinterpreting peaks and lows.

The real-time analysis that Pool Advisor provides is based on what they have access to, namely the DB2 system statistics data, and this is what is averaged. The Pool Advisor does not have or use any detailed object access information. Note, too, their collector runs all the time, requires multiple address spaces, and clients often complain about the overhead of their collector. To its credit, the Pool Advisor collects GETPAGE activity without a DB2 trace, but it is at an overall summary level .... and without detail trace data, they have no basis for any prediction of the effect of changes.

The Buffer Pool Tool (Responsive Systems) The Buffer Pool Tool is a new generation product that allows a performance analyst to evaluate the System and Application effect of changes to DB2 buffer pools in *Simulation* mode - without having to actually make the changes to a production system.

Simulations *predict* the effect of:

- Changing Pool Sizes
- Changing Pool Thresholds

- ❖ Moving Objects into Different Pools
- ❖ Moving Objects into New Pools that do not Currently Exist

Simulation / Prediction allows a much faster evaluation of tuning options than a trial and error methodology, and avoids costly mistakes with the production system. Since the Simulation shows the impact and the interaction of all objects together within a pool, is much more effective and accurate than attempting a pool isolation methodology for tuning. Quite directly, this is the only way you can get it right.

Temporary pool isolation, and performance evaluation of an object in one pool might take you the better part of a lifetime to complete for a large DB2 environment, and the fallacy of pool isolation is the lack of interaction between multiple or many objects. This cannot be determined or evaluated using isolation techniques.

Statistical Analysis provides a level of information for the overall System, each Pool, and every Object, that is simply not available from any other product. Additionally, a PC-based graphics component makes it easy to find the objects that are performing the best, those that are performing poorly, the impact each is having on its current pool, and which objects will obtain the greatest improvement from changes.

The PC-based graphic analysis component processes summarized data from the mainframe facilities, and shows the best and poorest performing pools and objects. An *analysis* of the current environment is given and also provides *tuning suggestions* for your system. These suggestions may encompass anything from changing pool thresholds, increasing pool sizes, and moving objects into different pools.

- Most importantly, Buffer Pool Tool predicts the I/O rate/second, the primary performance and tuning metric

Final Thoughts and Summary

We have seen how averages can be poor predictors of performance, price, etc. The selection of an interval greatly affects the accuracy of a prediction based on average. Over a long interval, average acts as a summary of past performance … it lags behind recent performance when using a moving average. Even by giving more weight to the recent past using an exponential moving average, averages still have inherent flaws when it comes to be accurate with respect to prediction. There is a great wealth of evidence from numerous fields that simulation is the only reasonable prediction technique to apply, especially when a computer is present. Simulation allows many possible scenarios to be evaluated, which lead to identifying peak conditions before they actually happen.