Everyday Benefits of Understanding Variability

K. Larry Weldon
Simon Fraser University
weldon@sfu.ca

Abstract:

Statistics professionals usually focus on the methods of formal inference. Informal inference is left to investment analysts, sport commentators, government bureaucrats and others who may rely on their intuition for guidance rather than a formal education in statistics. In this paper, I provide some examples of often-overlooked statistical phenomena that would be useful for the layman. The contexts for these examples are investment, sport, academic research, health, and lotteries. I suggest that statistics professionals should allot some energy to communicating such examples to the general public.

Many of us are involved in applying statistics to formal scientific or commercial research – either as teachers, researchers, or students. The emphasis in these contexts is on providing convincing evidence that the information we see in our data is reproducible, using widely accepted data analytic methods. However the utility of our discipline is actually much broader than formal research, and there is an opportunity to gain kudos from the general public by making this larger utility more visible. In this paper, I will suggest a few examples to illustrate this opportunity. I think this is important even for those whose main interest is formal research and graduate education, since it is primarily the general public that supports both research and education, politically and financially.

The following sections describe examples of opportunities in sport, investment, research and lotteries, for informing non-statisticians of the real-world utility of an understanding of statistics phenomena.

1. Sport – the quality illusion. Consider a typical playoff series of seven games, and suppose the A team wins the first three games. Most observers will, at this stage in the series, have the firm opinion that the A team is better than the B team and has a better than 50% chance of winning the fourth game. They may be right. But consider the situation in which both teams have equal "quality" in the sense that each team has the same chance of winning any of the seven games, that is a 50% chance (ties not allowed in a playoff series). For one team to win 3 games in a row, an event with probability ¼ must occur, and so this null hypothesis of equal-team-quality deserves some respect. Winning three games in a row is very weak evidence of superiority. The failure to recognize this might be called the "illusion of quality".

How is this insight useful to the layman? If the illusion is strong enough, gamblers will be willing to pay premiums for successful bets on the underdog. Suppose an enthusiastic fan of the A team agrees to pay 2 to 1 to the underdog bettor if the B team wins game 4.

That is, the A team fan risks $100 whereas the B team fan risks $50. If the teams are equal quality, the B team bettor has a winning strategy (over several such bets), since the average gain to the B team bettor in such bets is +$25. Backing the underdog can be a profitable strategy.

Of course it is not necessary to gamble to gain respect in this situation. Predicting a win for the underdog and winning almost 50% of the time will impress most skeptics!

2. Sport – when variability is the key

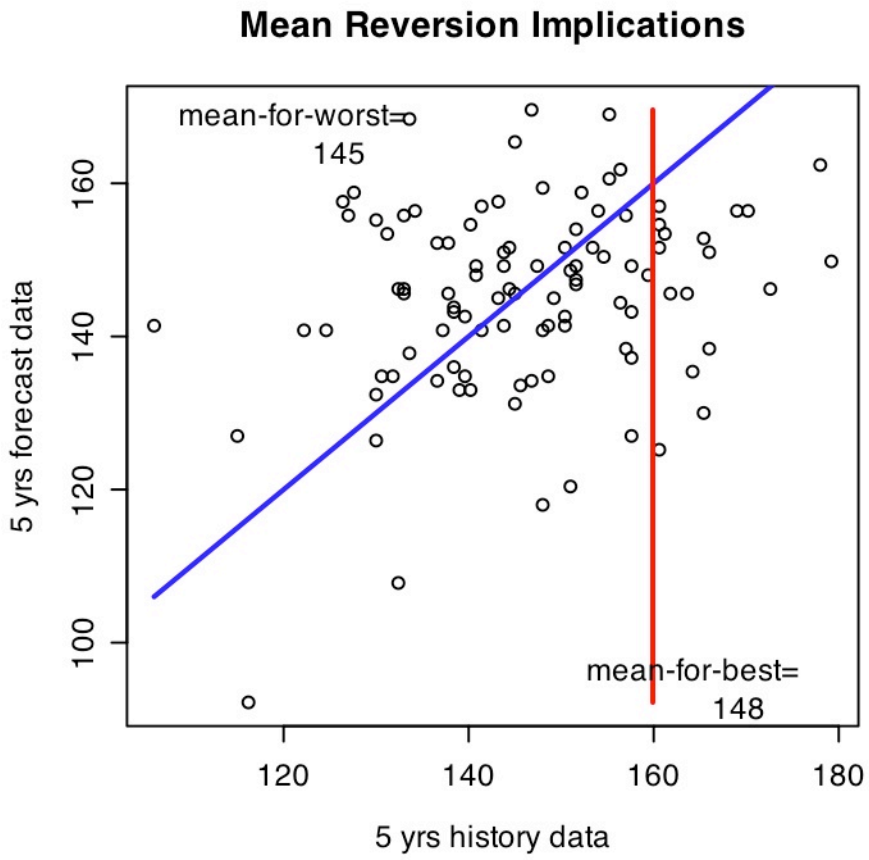
Golf courses typically have 18 holes with pars on individual holes varying from 3 to 5. It is common for golfers to think of the par 5 holes as the most difficult – perhaps the chance of exceeding par is greatest for these, or is perceived to be so. However, a case can be made for judging the importance of the holes by considering the variability of golfers' scores on those holes. Imagine a hole that, for the lucky few, has a par 3, but for which there is a considerable chance of losing one's ball in a water hazard, or in the adjoining woods! This is clearly as more important hole than a long par 5 with few hazards. It is a hole like this one that separates the winners from the losers, over an eighteen-hole round. This example, and many others, are described with data in Clarke(2007). It is a nice example of a situation in which the variability is of interest in itself, not merely as an adjunct to the mean.

The general theme in Clarke's paper is that variability of performance is as important a parameter of performance as average performance. He explores the setting of record performances, in which variability plays an important role. The lesson here is that every performance in most sports is a combination of competence and luck, and we need to appreciate the role of luck in assessing and predicting performance.

3. Investment – back-the-winner fallacy

Non-professionals approaching investments of life savings either have to take the advice of a professional, or try themselves to seek a profitable and reliable investment strategy. Aligning the interests of the professional and the client is not so easy to do. One strategy that the amateur may use is to look to mutual funds. Even small investments can be diversified so that the probability of loss is reduced compared to investment in a small number of stocks or bonds. However, the long list of mutual funds is not easily ranked except by past performance. The merit of ranking by past performance in predicting future performance depends on the mix of skill and luck for the management of each fund, and the evidence shows that luck is by far the larger portion. By "luck" here we mean the occurrence of events effecting financial markets whose prediction or timing were unknown to the managers in advance of the events. If luck is a large portion of the determinant of future investment performance, then past performance is an unreliable predictor of future performance. We can demonstrate this with a simulation calibrated to match past market activity.

It turns out that a fund whose daily valuations have increases 55% of the time, and with a mean absolute size of increase or decrease of about 0.3% of current value, does mimic stock index history, at least in Canada. The differential between a good fund and a bad one is very small, and this 55% might only vary over 54%-56%. This small difference applied over a five year period would tend to result in annual returns of 8.5% for the 54%, and 12.6% for the 56%. In the simulation described below, we simulate the daily experience of 100 funds with a quality index of 55% ± 1%, over a five-year period, selected independently for each of the 100 funds. Then we choose the best 15 of these outcomes and, using the same quality funds as were originally generated for each of the 15, simulate the next five years of data. The graph shows a typical outcome: the 15 best from the historical data did not achieve significantly better returns than the 84 worst from the historical data. The reason for this surprising result is that the variability due to the accumulation of many small changes swamps the trend induced by the small probability of market increase.



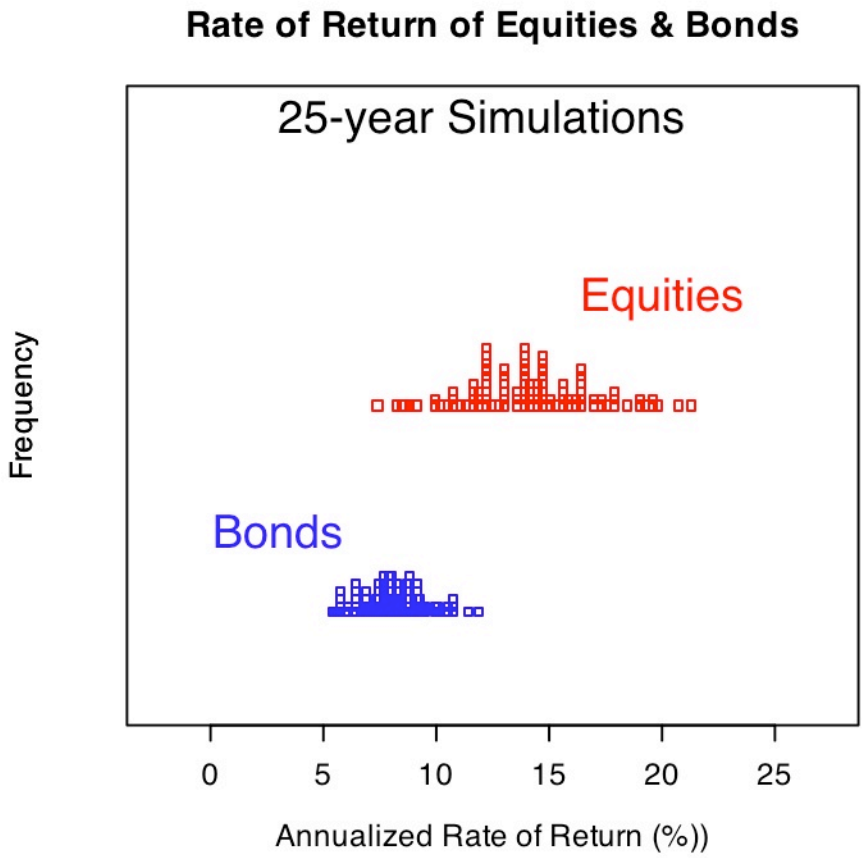
The lesson in this is that historical performance as measured by annualized returns is an almost useless indicator of future performance. There may be information that recommends one fund over another, such as the style of the fund in relation to economic forecasts, or the country mix, or changes in management personnel, but past performance is not a helpful indicator. "Backing the winner" does not work well in this context.

Note that the result does not require the simulation demonstration to be convincing. There is obviously a component of unpredictability in the stock market, and mutual funds would be expected to differ in performance even if all fund managers were equally competent. However, the magnitude of the effect of this unpredictability is much larger than most people expect, and this is the message that the statisticians can convey to the general public.

4. Investment – variability vs risk.

When the market value of a company is variable, investment in the company is sometimes described as high risk for the investor. If the investor has only one company in the investment portfolio, then this is a reasonable description: there is a real risk that the company could lose value indefinitely or even become bankrupt. But, if the portfolio includes a large number of companies, or if it is a single mutual fund with a large number of companies, then variability is less directly related to "risk". The investor is only likely to lose money if there is a short time horizon – in other words if the investment must be liquidated at a time when the market valuation might be low. For the longer time horizon, more variable investments tend to produce higher capital appreciation. Small companies make greater percentage returns, in aggregate, than larger companies, and company stocks generally have higher returns than bonds. In other words, the investments generally considered to be "low risk" (meaning low short-term variability). such as large companies and bonds, will usually have lower returns as well, especially in the longer term.

An example of a long-term investor is a pension plan member. Contributions are made to the plan over 25-45 years and withdrawals are made over an additional 10-25 year period. In this situation short-term variability is irrelevant and long-term return is very important. Many investment advisors suggest that an investment portfolio, even in this long term situation, should have about 40% bonds (low variability, low return) and 60% equities (high variability, high return). For the long-term investor, this "balanced" portfolio almost guarantees **under**performance compared to the 100% equity portfolio. A simulation illustrates the impact of the small equity advantage over a 25 year period – see the the figure below. This demo uses daily increase probabilities of .544 for bonds and .547 for equities, which are based on actual index data over a 50-year period in Canada.



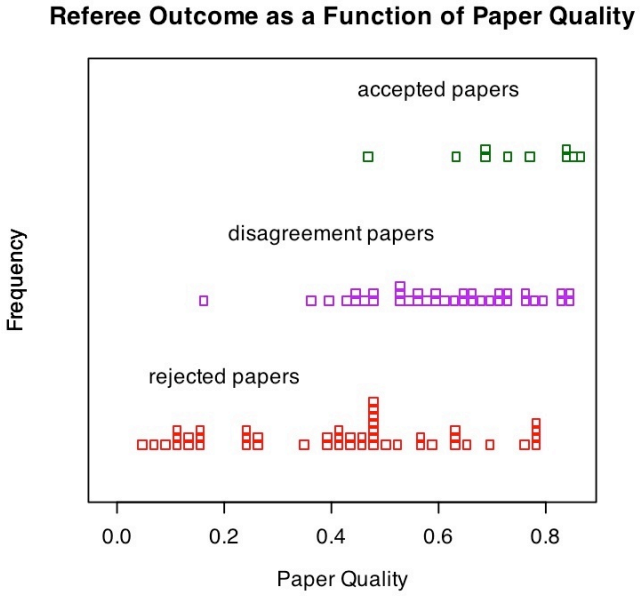
The implication from this example is that it is the time horizon, rather than the tolerance for variability, that should be the main determinant of the investment class mix for an individual's investment portfolio. The message from stock brokers and mutual fund salespersons usually concerns the degree of risk as measured by short-term variability. For the pension (or other long-term) investor, this is the wrong criterion. Statisticians understand variability as different from "risk" in its everyday meaning, and we should inform the general public about this distinction.

5. Research – the randomness of peer review

While "research" is not really a common activity of the "general public", it is certainly pursued by many professionals who are unfamiliar with statistical theory. In a sense this group is "general public" from the point of view of the academic statistician. The example used in this case is the peer review system. It actually has many parallels in assessment procedures outside of formal research – for example, in procedures of the courts, immigration, or admission to education institutions. So the peer-review system in journal publication procedures is a suitable model for a wide range of peer-review systems.

A common procedure in peer review systems for academic journal submissions is for two members of a list of "peers" to be appointed for review of a particular paper. If both peer referees give positive reviews, the paper is accepted for publication. If the referees are not in agreement, a third reviewer is appointed to break the tie, one way or the other. This method works well if all reviewers have the same standard for providing a positive recommendation. But this ideal is hard to achieve. A realistic situation would have the proportion of positive recommendations that a referee provides on a certain batch of articles to be anywhere from 10% to 50%, depending on the referee. In this case, the outcome of the review process for a particular paper is largely a function of the choice of reviewers, and not so much determined by the merit of the submission.

The following graph shows the result of a simulation of this process. We assume a wide range of article qualities and a wide range of reviewer tendency to approve. For a given paper-reviewer combination, the probability of a positive recommendation is dragged toward 0 or 1, from the referee's average tendency, depending on the quality of the paper. However, as the simulation demonstrates, the variability of reviewer tendencies still causes good quality papers to be rejected, and mediocre quality papers to be accepted. If we focus on the papers whose quality is among the top quartile in quality (in the simulation reported in the graph below), 9 would be accepted and 16 would be rejected. So many good papers had an unlucky result. The point of the simulation is that the peer review process is not as reliable as it is often assumed to be. Of course, the parametrization used can be criticized, but it would take quite unrealistic changes to eliminate the qualitative result shown.



6. Health – mimicking the natural environment

Although lifestyle experts have a wide range of prescriptions for a long and healthy life, much less has been written about the implications of our evolutionary past for such prescriptions. The modern wild animal world gives us some hints of the natural environment through which our ancestors evolved into modern homo sapiens. It would seem to be quite likely that this ancient environment posed many challenges for survival, including shortages of food, vagaries of weather, and scourges of disease. It may be an over-simplification to say that we evolved to survive changing conditions, but if so, it might be that constant conditions are not optimal for us. One very small bit of evidence in this direction is given in the paper by Brewster et al (2005) . Apparently patients requiring artificial ventilation because of impaired lung function do better if the rate of supply of oxygen to their lungs includes a random component over time. A rather large leap of inference might suggest that we consider that a variable environment might be more healthful than one that is strictly controlled. To think of variability as a positive factor rather than as "error" is something statisticians should make more of in communicating with the public!

7. Lotteries – expectation and hope

Lotteries raise huge amounts of money for their owners (often governments), and the reason for their popularity is that it makes the fulfillment of a fantasy a possibility. Of course, no money would be made if the average return of a lottery ticket exceeded its cost. So while the possibility of winning a major jackpot looms large in the mind of the ticket purchaser, the probability of it is miniscule. In a typical legal lottery, the chance of winning a major jackpot is so small that a lifetime of participation will almost never include a jackpot win. In fact, if we consider a devoted fan of the lottery who purchases 100 tickets every week for 60 years, the chance of missing out on the jackpot is 98 percent! Or, if they use the alternative desperate strategy of buying all the tickets in one particular lottery, they will get approximately one-half of their investment returned. There are two lessons here: One is that a one dollar ticket is worth and average of fifty cents after it is purchased (until the outcome is announced), and the other is that the hope of winning should be a faint hope. The cash flow in a lottery is a simple thing to understand. It is surprising that so many people consider these public lotteries as a reasonable investment.

Are there other vehicles for hope with better expectations? The stock market, possibly diversified through a mutual fund, may actually provide a way to improve on the prospects of public lotteries. Of course, if the public lottery is viewed as a form of charitable giving, then it provides a convenient way to do so.

Summary

Professional statisticians (university instructors, consultants, official statisticians) are trained to provide specialized services to specialized recipients. This training includes certain ideas that have implications for the general public. The discipline of

statistics would be more widely appreciated if professional statisticians made more use of the opportunity to expose their discipline to the general public. The fields of sports, investment, peer review, health and lotteries are examples of areas that statisticians can use in this "public relations" effort.

References:

Brewster, J.F., Graham, M.R., and Mutch,W.A.C (2005) Convexity, Jensen's inequality and benefits of noisy mechanical ventilation. J. R. Soc. Interface 2: 393-396. Online Journal of the Royal Society.

Clarke, S.R. (2007) Raising Interest in Statistics through Sporting Predictions one the Internet. Invited paper to be presented at the ISI Conference in Lisbon, August, 2007.