The Hyena's Corruption of Data Integrity

David Albrice, May 18, 2015

"Statistics are used much like a drunk uses a lamppost: for support, not illumination" – Vin Scully.

Once when an enormous Lion was asleep, a little Mouse eagerly began running beside him. The Lion startled awake, immediately placed his huge paw upon the Mouse, and opened his jaws ready to swallow her whole.

"Pardon me, O' King!" cried Mouse, "I did not mean to disturb you. I am just a lowly messenger trying to deliver some important asset management news. Please don't shoot-the-messenger, as they say."

"Asset management?" growled Lion. "What would such a small creature as you know that would be relevant to such a large and ferocious beast as I?"

"My message is important, but before I deliver it, please promise that you shall not kill me. I will be most grateful and perhaps one day I may be able to do you a favour in return."

Lion was so tickled at the audacity of Mouse that he lifted up his huge paw, listened to her important message, and then let her go.

Periodically Mouse would return to deliver additional messages to Lion.

But one day... Lion was caught in a trap by local hunters. He could not free himself and was nervous that they would return for their trophy or that Hyenas would kill him first. Just then Mouse happened to pass by, and seeing the sad plight of the Lion, she immediately gnawed away the ropes that bound the King of the Beasts.

"Was I not right?" said Mouse, "Are we not kindred spirits?"

"Yes," Lion replied, "We are kindred spirits. Do you have another message for me today?"

"I do," said Mouse, "but first we need to get out of here. The Hyenas are coming and they may corrupt our professional collaboration. Let's return to your den first for a discussion on **integrity** and then to my burrow for a look at **credibility**."

1. The Lion and his Integrity (Statistical Data)

Lion and Mouse resumed their conversation once back at the Lion's den. "I have a ravenous appetite and I consume massive quantities of data," said Lion. "I need lots of big data for my power and strength."

Lion went on to explain: "Just as statistical science plays a fundamental role in both the insurance and finance industries, so too is it essential in asset management. It is particularly useful in making forecasts on the service life of assets, such as roofs and boilers."

Lion gave some examples:

- The ***average*** life of a 2-ply SBS roof is 25 years.

- Based on a ***statistical population*** of 2,000 roofs, we know that certain kinds of metal roofs can last for 40 years.

- The longest surviving ***statistical outlier*** for a BUR type roof is 38 years

- A ***representative sample*** of assets in a study on the lifespan of asphalt shingle roofs

"The insurance industry uses a variety of tools to navigate inside the data jungle and survive their hyenas. Its actuaries calculate risks with elegant models and sophisticated techniques," said Lion. "Here are a few of the most important tools that are relevant to asset management." He handed a list to Mouse.

- ***Survivor Curves*** indicate the number of assets that are retired from service at different ages.

- ***Scatter Plots*** reveal correlations between two or more attributes of an asset, such as its age and its condition.

- ***Regression Analysis*** fit predictive curves to data points.

Lion then handed Mouse a chart to illustrate two examples of asset survivor curves. "These probability distribution graphs allow statisticians to analyze mean, median, mode, dispersion, standard deviation and other stuff in statistical science. But data in the real world is a little messy and usually asymmetrical. Take a look at this chart to see what I mean."
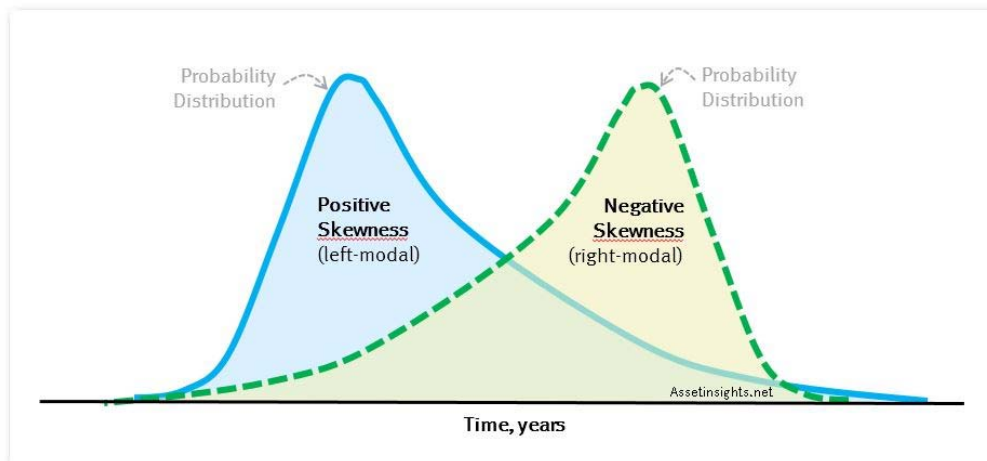
"I understand." said Mouse. "In my burrow I have copies of some of the published service life tables developed on statistical principles by organizations such as BOMA, ASHRAE, RSMeans, and Whitestone. I use these life expectancy tables in my daily work. But I am careful not to rely too heavily on these averages."

Mouse continued, "As the statistical population gets larger, the shape of the curves change to reflect new data. This gradually improves the quality of the forecasting and management of the risk."

"But statistics alone are not always enough," said Lion. "The all-powerful insurance industry recognizes that its enormous statistical data needs to be tempered with empirical data. Let's take the example of a life insurance policy, the underwriters will first look at the age, gender, geographical location, smoking habits and other hard data on a person to determine average lifespans (statistical) but this will need to be aligned with the results of blood and urine tests (empirical) in order to better understand the risk," said Lion.

"Since we are now getting into discussions about data credibility, we should move to my burrow" said Mouse.

2. The Mouse and her Credibility (Empirical Data)

"I nibble my food in small pieces," said Mouse. "I focus more on qualitative data rather than quantitative, and this requires a very different approach to yours, Lion."

"I'm a technical expert, an engineer, who analyzes local empirical data on the condition and performance of assets. I do this in different parts of the world, in different climate zones, under different operating conditions, and in different service environments."

"My army of technical experts is constantly identifying discrete distress metrics and other factors that affect the durability and longevity of assets in subtle ways." Mouse handed Lion a list to illustrate her point:

- There is ***blistering*** in the southeast corner of the roof at building #214

- There is ***corrosion*** on the cap flashing at building #19

- There is ***cracking*** of the concrete on the 2nd floor at building #2

- There is ***staining*** of the facade on east elevation at building #25

"My engineers use a variety of tools to cut through the thick tangled undergrowth in the jungle with precision," said Mouse, as she pulled out another list and handed it to Lion.

- ***Non-Invasive Diagnostics*** reveal concealed conditions through the application of sophisticated technologies such as thermography, ferrography, smoke testing, water testing, impact echo, ultrasound, vibration analysis

- ***Destructive Testing*** reveals concealed conditions through exploratory openings

- ***Performance Curves*** indicate the gradual loss of function of an asset with intervals of various forms of maintenance and tests to elicit empirical data on the remaining life of an asset

- ***Degradation Curves*** indicate the rate at which an asset is experiencing physical deterioration

"Here's an example of a degradation curve that shows the different stages of physical deterioration and the increasing risk as an asset approaches functional failure," said Mouse.
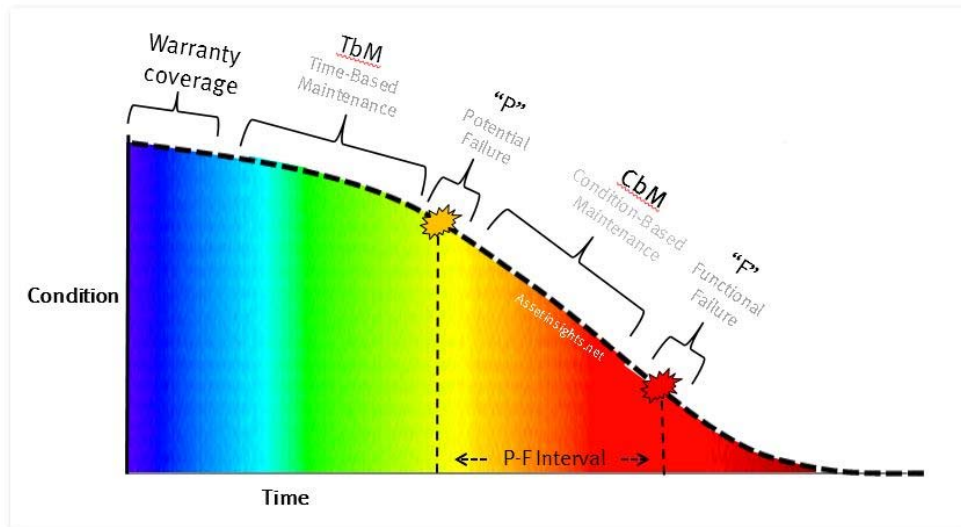
"I know that empirical data is necessary but not always sufficient on its own," said Mouse. "Just as the insurance industry tempers its powerful statistical data with empirical data, so too must my empirical data be reconciled with your larger framework, Lion, so that it has context and meaning."

"My engineers must be rigorous in their data collection protocols or they will compromise your statistical models," said Mouse. "And..." said Lion, "my statistical models give your empirical data the essential structure for classification and analysis."

Just then, Lion and Mouse heard a rustling outside the burrow. It was Hyena on the prowl. "I heard you two data geeks talking, and I wanted to join in your conversation" said Hyena. "After all, there are more Hyenas than Lions in this jungle" he smirked.

3. The Hyena and his Corruption (Bad Data)

"I am a scavenger that eats absolutely anything," snarled the Hyena. "I don't have a delicate, discerning palate like you, Lion and Mouse. I care nothing for qualitative or quantitative data. I don't distinguish statistical data from empirical data. It all gets ground up and digested in my stomach anyway."

Hyena looked at his list and a wry smile appeared across his face.

- The age of the asset is unknown and there are no maintenance logbooks due to *poor recordkeeping*

- The data on the asset is erroneous due to *improper methodology* applied to the data sampling, data collection and data analysis.

- If the data do not align with the *vested interests* of certain stakeholder, they are adjusted to match accordingly.

- A *sampling bias* is permitted in the data set so as to return the results that were being looked for by those with a vested interest.

"Here is an example of the erroneous data that I find so tasty," said Hyena. "Look at just how clean and neat my data are..."
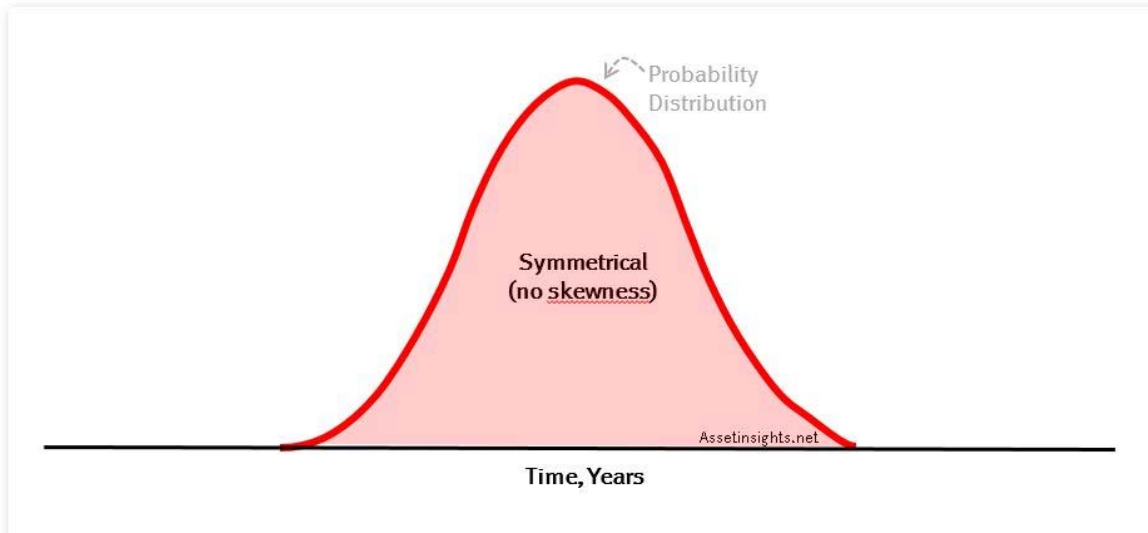
"Who cares for quality when there is so much quantity. Who cares for nuance and subtlety when everything is edible," snarled the Hyena.

4. Lion and Mouse forge ahead

"Hyena, there is no place for you here," roared Lion.

"We have what we need for credibility and integrity," added Mouse.

"We have worked diligently to ensure that you, Hyena, do not corrupt our asset management models. We have adhered to rigorous data collection methodology and have disclosed all our assumptions. We have not manipulated our data to make it fit some preconceived notions."

"I want to thank you kind Mouse," said Lion. "I am truly indebted to you. My statistical data rests on your empirical data as its foundation. With your help, I avoid Hyena's attempt at Garbage-in-Garbage-out."

"I too am indebted to you, Lion," replied the Mouse. "My empirical data needs a framework in order to derive its meaning. With your help, I avoid Hyena's attempt at disrupting our Single-Point-of-Input and Single-Point-of-Control."

"We-are-kindred-spirits... We-are-kindred-spirits...." sang the Lion and the Mouse - in chorus - as they glanced over each other's shoulders for any signs of the Hyena.

What are your strategies for keeping data safe from Hyenas?

*****

This blog is loosely based on the Aesop's Fable The Lion and the Mouse.

For those interested in deeper study, a paper on the role of empirical science and statistical science in asset management can be downloaded here.

David Albrice is a Senior Asset Management Specialist at RDH Building Engineering Ltd. His experience with lions comes from the years he lived in South Africa. David can be followed on [Twitter](#) and additional asset management articles can be read on [LinkedIn](#).