Layers of Meaning:
Free Statistics Textbook

by M. E. Kabay, PhD, CISSP-ISSMP
Professor of Computer Information Systems
School of Business & Management
Norwich University, Northfield VT

Why should information assurance (IA) professionals care about applied statistics?

One reason is that distinguishing between chance variations and unusual anomalies can lead system, network and security administrators to problems that might be discovered much later, when they approach disaster status. In “Pay Attention to Anomalies,” I pointed out the operational consequences of a software failure that ignored interruption of automatic payments by members of a health club. In the article, I showed a familiar diagram illustrating the *change in slope* or *inflection point* in a growth curve as a possible indication that something has changed in the operational processes about which we are concerned.

But how do we tell if something is *significant* or merely a minor chance variation?

Decision under uncertainty is one of the descriptions of applied statistics, although it’s hard to know who said that first. Making decisions when the facts don’t allow a clear-cut decision among alternatives is difficult; it’s easy to decide to cross a street when there are no cars visible at all, but the moment cars are in sight there are questions of how fast they are going, whether pedestrians are *likely* to be visible, whether the drivers are *likely* to be insane sociopaths who want to run over pedestrians, and so on. Admittedly, one could push the example to ridiculous extremes, such as arguing that we have to weigh how *likely* it is that the street will cave in just as pedestrians step out onto the pavement, whether a meteor is *more likely* to strike a pedestrian crossing the street than a pedestrian standing on the sidewalk, and so on.

All of these questions imply that absolute certainty is impossible; we are victims of randomness and have to cope with uncertainty as best we can. Look at all the cases where italics highlight “likely;” we take uncertainty for granted in the real world and generally cope with it intuitively, without having to use statistical methods explicitly or consciously.

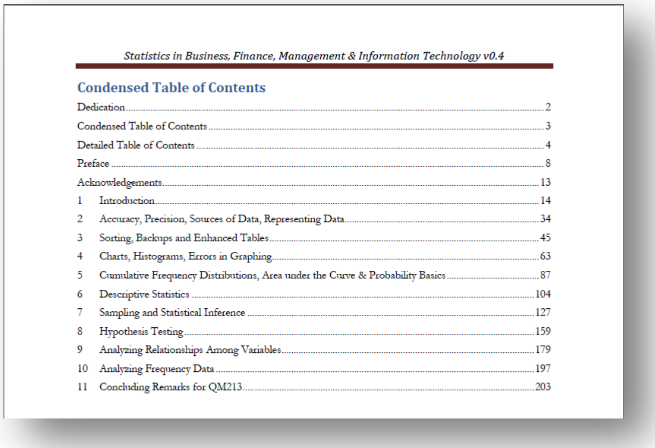
Applied statistics includes an area called *hypothesis testing* that deals specifically with evaluating the *likelihood* that a particular model of external reality – that a particular model of external reality – the *hypothesis* – is *consistent* with *unbiased* observations. For example, turning back to network security, we might say that we hypothesize that the rate of attacks on our gateway security devices today is consistent – not from expectation – with recent attack rates. So if we expect 200,000 attacks a day on our IP addresses and today we see 300,000, does that increase mean that someone new is attacking our systems, or that old adversaries are focusing on us – or is it just the luck of the draw?

Answering the question is beyond the scope of this column, but intuitively, a lot depends on how *variable* our historical data are. For example, if the number of attacks ranges from, say, 150,000 to 250,000 over the last month, a value of 300,000 might not be a clear indication of a real change. But if the attack rates vary from 190,000 to 210,000 for the last 30 days, maybe a value of 300,000 is something to make admins' eyebrows go up.

This column is not intended to teach anyone details of statistical methodology: it is a pointer to a new version of a statistics textbook I've been working on since 2010. Now in its fourth pre-publication version ("v0.4"), *Statistics in Business, Finance, Management and Information Technology: A Layered Introduction with Excel*< http://www.mekabay.com/courses/academic/norwich/qm213/statistics_text.pdf > is a 200-page introduction to practical applications of statistics for estimation and hypothesis testing. The material is the basis for my QM213 "Business & Economic Statistics I" course< http://www.mekabay.com/courses/academic/norwich/qm213/index.htm > at Norwich University< http://www.norwich.edu > and is freely available to all users. My only requests are that no one post copies online (in public or on intranets); simply point to the file instead – that way when I make corrections, everyone will have access to the latest version. And naturally, I would be very cross indeed if anyone *sells* what I *give away free*!

The text provides practical guidance on using Excel 2010 for routine statistical processing; more important, I put a lot of work into explaining the underlying reasoning for all these methods. This is a book for practitioners and should allow users to improve their understanding of *which* statistical methods to use *when*.

The short-form table of contents will give readers a sense of the topic areas.



Statistics in Business, Finance, Management & Information Technology v0.4

Condensed Table of Contents

The structure of the text is unusual in that topics are introduced several times, helping readers and students get used to *why* various methods are used before they are introduced to *how* to carry out the computations.

I will be grateful to readers who point out errors or possible improvements in the text – this project is definitely run under the standards of continuous process improvement.

Have fun (assuming *having fun* and *statistics* are not mutually exclusive for you).

* * *

For a primer about applied statistics aimed at IA professionals, see "Understanding Computer Crime Studies and Statistics, v6" which became chapter 10 in Bosworth, S., M. E. Kabay, & E. Whyne (2009), eds. *Computer Security Handbook, 5th Edition*. Wiley (ISBN 978-0471716525). 2 volumes, 2040 pp. AMAZON < http://www.amazon.com/Computer-Security-Handbook-Volume-Set/dp/0471716529/ >

* * *

M. E. Kabay,< mailto:mekabay@gmail.com > PhD, CISSP-ISSMP, specializes in security and operations management consulting services and teaching. He Professor of Computer Information

Systems in the School of Business and Management at Norwich University. Visit his Website for white papers and course materials.< http://www.mekabay.com/ >

* * *