A study sponsored by Cisco Systems and carried out by InsightExpress using responses from more than 2,000 respondents in 10 countries indicated that accidental and deliberate violations of security by insiders are a serious threat to data confidentiality.

Jim Duffy of Network World wrote, "Thirty-nine percent of IT officials surveyed perceive negligence among employees as the main reason for the data security risk, while one in five pointed to disgruntled workers as the source. One in three IT respondents said portable hard drive devices are their top concern for how data is leaked -- more than e-mail (25%), lost or stolen devices (19%) and verbal communication with non-employees (8%)."<http://www.networkworld.com/news/2008/111208-cisco-study-internal-security.html>

He added, "One in 10 employees surveyed admitted stealing data or corporate devices, selling them for a profit, or knowing fellow employees who did. This finding was most prevalent in France, where 21% of employees admitted knowledge of this behavior."

I’m sorry that the report (not Jim Duffy!) included arrant nonsense. Let’s look at some simple elements of statistical analysis (no, really simple: not even one formula today).

The problem with the statement about admitting stealing or knowing employees who stole is that it combines different causative factors that can result in the response. For example, suppose we are studying the effect of a new series of security-awareness cartoons on employees. One could form two groups, the cartoonified group (C+) and the uncartoonified group (C-) and then study their susceptibility to, say, phishing attacks sent to them via e-mail. Sounds great! We do the test and end up with the following table:

<table>
<thead>
<tr>
<th></th>
<th>Tricked</th>
<th>Not Tricked</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-</td>
<td>72</td>
<td>128</td>
</tr>
<tr>
<td>C+</td>
<td>52</td>
<td>148</td>
</tr>
</tbody>
</table>

For statistics aficionados, we compute a chi-square statistical test of independence with a value of 4.219 (with 1 degree of freedom) for a probability of 0.04 that there is no relationship between cartoon exposure and resistance to phishing. So obviously exposure to the cartoons increased resistance to phishing messages, at least at the 0.05 level of significance, right?

Ah, but suppose that, without reporting the fact, we actually have an additional orthogonal (independent) factor defining two groups of employees: those who have previously been given a full-day security-awareness workshop (W+) and those who have not (W-). Well, that means that there are actually four test groups: W-C-, W-C+, W+C- and W+C+. And then we find out belatedly that the results, when classified with the additional information about security training, are as follows:
So the results with both variables displayed indicate quite a different story: the cartoons have very little effect on people who had no security-awareness training but there was a noteworthy improvement after exposure to the cartoons among those who had been trained. In statistical terms, we call this phenomenon an interaction between the independent variables (workshops and cartoons); there are tests for decomposing the effects precisely (the log-likelihood ratio, G, is my favorite). For readers who have studied analysis of variance (ANOVA), the G-test is the non-parametric equivalent of a multivariate ANOVA. But enough of this airy persiflage.

In statistical analysis, we refer to confounded variables when an analysis varies more than one attribute in a measured variable (an independent variable) and then ascribes fluctuations in a result (a dependent variable) to only one of the variables. In our example, the study confounds exposure to cartoons (what the survey claims accounts for resistance to phishing) with the unreported variable, exposure to security-awareness training.

You can see that because the original study confounded the two variables (cartoon exposure and awareness training), the analysis of the pooled data was misleading: it falsely ascribed the difference in response to phishing to the cartoons alone. So now let’s go back to the issue of people who admitted to having stolen data or to knowing someone who did.

The general principle here is as follows: Statements of the form "X% of respondents admitted doing Y or knowing of coworkers who did Y" don’t mean anything. They confound a number of factors into a meaningless jumble:

1) How many people do Y?
2) How many people who do Y are willing to admit it to an interviewer or on a survey?
3) How secretive are people who do Y about letting coworkers know about their actions?
4) How many people learn about a single person's transgressions?

I won’t even discuss the possibility that some people will report personally-held beliefs or rumors they have heard.

The issue of responsiveness (factor 2) is inherent in all studies and surveys, but factors 3 and 4 are at the heart of the problem here. If criminals are blabbermouths, those “knowing of coworkers who do” will rise; similarly, if the social networking of criminals is high, more people will know about the crimes than if the criminals are relatively private people. In any case, the confounding of doing the crimes and knowing about the crimes makes the statistic useless.

As a simple example of how misleading the garbled statistic can be, imagine a reduction to absurdity. Suppose a single person in a company of 10,000 steals trade information and gets arrested and convicted. The security department releases an alert about the case as part of the security-awareness program and all 9,999 other employees therefore know about the case. An interviewer arrives some time later and interviews a hundred employees, all of whom say that they have NEVER stolen trade secrets but ALL of whom say they know of someone who did.
The report would state that “100% of the employees admitted stealing data or knowing fellow employees who did.”

Bah, humbug.

* * *


* * *

M. E. Kabay, PhD, CISSP-ISSMP <mailto:mekabay@gmail.com> specializes in security and operations management consulting services. CV online.<http://www.mekabay.com/cv/>

Copyright © 2009 M. E. Kabay. All rights reserved.

Permission is hereby granted to Network World to distribute this article at will, to post it without limit on any Web site, and to republish it in any way they see fit.