Finding practical applications for significant results

Introduction – Confusing "significant" and "important"

In the management research literature, it is important to understand that in the term "statistically significant," the word "significant" doesn't mean the same thing as "important." When a correlation result from a regression analysis is described as "statistically significant", it means that the probablility that the result is an accident of random variation has been precisely determined, and that the chances are lower than some defined minimum.

That defined minimum is identified by "alpha;", also called the "confidence level." E.g., a confidence level of 95% is the same as = 0.05, or "p<.05", and means there is a 5% chance of error. So, a statistically significant result, at a 95% confidence level, has a 5% or less probablity of being incorrect.

For a regression analysis, reporting that a result is "statistically significant" means that the proposed correlation can be shown to exist in the data.

Statistical significance tests determine "whether the phenomena that researchers observe are caused by sampling error (accidentally picking unrepresentative subjects) or 'real' effects" (McCloskey,1995). For a regression analysis, reporting that a result is "statistically significant" means that the proposed correlation can be shown to exist in the data. However, it says nothing about how important that correlation is. A correlation which is statistically significant is not necessarily significant in importance.

> Concern about the statistical significance of effects (whether they exist at all) has tended to preempt attention to their magnitude. … Although not unrelated, the size and statistical significance of effects are logically independent features of data from samples. Yet many research reports, at least implicitly, confuse the issues of size and statistical significance, using the latter as if it meant the former. (Cohen and Cohen, 1975, p6)

However, another output from a regression analysis, the magnitude of the effect size for the correlation, can often be indicative of that correlation's importance. And yet, in the papers surveyed by [**CITE**], 92% fail to discuss the magnitude of the coefficients at all. Instead, 93% used statistical significance as the only criterion for the importance of a result. And, unsurprisingly, in 97% of the papers, the conclusions are also solely based on statistical significance.

In nearly all of the papers in that survey—98%—the term "significance" is used in an ambiguous manner. No clear distinction is made between statistical significance and other types of significance. As Cohen and Cohen described above, in most cases statistical significance is implicitly used to mean "important".

Purpose of doc

If managers are going to be able to use relevant research from academia as part of the process of making strategic decisions, they need to be able to figure out what research actually is relevant. While all the empirical studies they may find in the management literature will report statistically significant results, they won't necessarily be useful for making management decisions. While the inputs and outputs analyzed by an empirical study may be relevant, that is not sufficient to establish that a study is relevant to a real-world decision. The way that researchers interpret and present the results may impede, if not completely obscure, the implications of those results for real-world decision-making. And the results themselves, while statistically significant, may in the end describe too small of an effect to be worth a manger's effort to pursue.

This article aims to address that issue—identifying the potential utility of academic research in practical application—from both directions, discussing both how to read research articles and evaluate their potential practical utility, and how to write research articles in a way that increases their potential practical utility.

Hypothetical example

How can the statistically valid results of an empirical study be interpreted and presented in such a way as to be of no use to a manager? Consider the following hypothetical example.

Let's assume that you are the manager of a plant that uses 100 gallons of pink paint every month for production. You pay $10 a gallon for paint, but your vendor makes a proposal: you can instead buy white paint and red paint for $8 a gallon, and blend the two into pink paint. You would save $200 per month.

But there would be costs from the change. Additional time, labor, training and equipment. You have an intern look into it, and it turns out that those cost add up to $100 per month.

However, there is an chance that something could go wrong with using two colors of paint blended together as part of the production process. The added complexity could potentially cause a non-conformance of the product, requiring reworks (if caught in-house) or replacements (if found by the customer). All of which can be reduced to a cost.

Your intern calculates that should the two-color process cause a non-conformance, it would cost $200 to remedy. In any month without a non-conformance you'd still save $100 with the two-color process, but in a month with a non-conformance (we'll assume there's never more than one), you'll lose $100 by using the two-color process.

What you need to know, is the likelihood of a non-conformance. If it's less than 50% per month, then you should use the two-color process. If it's more than 50% per month, you should not use the two-color process.

So, you hand the problem over to an academic researcher to whom you've given data in the past. She passes it on to a PhD candidate, and a year later you get your answer.

The researchers have gathered a lot of data, and run a regression analysis on it. The candidate tells you that, "We are 99-percent confident that using two colors of paint instead of only one will increase the probability of non-conformance in your product."

And after a long pause, you say, "Yes, but by how much?"

"The increase will be greater than zero percent," says the candidate. "We are 99-percent sure of it."

After another long pause, you say, "Can you do any better than that?"

And the candidate laughs. "Better than 99-percent? Oh no, I don't think so. That's a very significant result!"

That example may be cartoonish, but it illustrates a real problem: real-world managers don't get the answers they need from management research. Because, while the PhD candidate in the scenario was probably correct that the result of their analysis was "very significant" from a statistical point of view, that same result was of no use at all to a manager trying to make a specific decision.

If it had been unclear whether or not the two-color paint process would increase the probability of non-conformance, then a result of "greater than zero percent" might have been a valuable insight. But the questions facing managers of businesses are rarely about whether or not an effect exists. What managers need to know is the strength of the effect, so they can balance it against other effects when making their decisions.

This is why managers and researchers in management should be wary of null-hypothesis significance testing. NHST only looks for non-null effects, for effects with a magnitude greater than zero. What it doesn't do is distinguish very well between effects with minor or major influences on outcomes.

So, empirical studies that use regression analysis can find statistically significant correlations that have large, important effects on outcomes; and they can find statistically significant correlations that have tiny, trivial effects on outcomes; and they can also find statistically significant correlations at every level of importance between those two. The statistical significance of a result in no way assures that it is an important one.

Which is, like much of statistics, completely counterintuitive. How can a result be both "significant" and "trivial"? The rest of this paper will hopefully explain that, and therefore give some insight into how to critically evaluate empirical research, and how to decide if the "significant" result reported in that research is an important effect, or merely a trivial factoid.

Conventions of the doc

Because of the purpose and context… we'll use some specific terms in place of the general terms used in pure statistics. dependent →outcome, independent →input, …

Outline the doc

[**FIX** This order doesn't work with the text. Or maybe the text needs to be forced into this order?]

First we'll look at how a statistically significance result can be either important or trivial, and to further explain what it is that statistical significance measures, we'll discuss Type I and Type II errors, specifically in relation to correlation analyses.

Then, given our renewed understanding of statistical significance, we'll discuss why bigger isn't always better when it comes to data. One of the specific problems covered will be how using big data can obscure important results with trivial ones. A related issue is the failure of authors to interpret the results of regression analyses of big data, which increases the difficulty of finding the important results among the larger number of trivial ones.

After a brief section on why this distinction between important and trivial effects is particularly important to management studies, we will give some guidelines for both how to read and how to report the correlation analysis results of empirical studies in management.

"Significant" or "Important"

> The function of statistical tests is merely to answer: Is the variation great enough for us to place some confidence in the result; or, contrarily, may the latter be merely a happenstance of the specific sample on which the test was made? This question is interesting, but it is surely secondary, auxiliary, to the main question: Does the result show a relationship which is of substantive interest because of its nature and its magnitude? (Kish, 1959)

How can a significant result not be important?

"Statistical significance" is a mathematically pure characteristic of a correlation. Meaning that it doesn't require any interpretation, and that it is independent of the particular phenomena which the correlation's variables describe.

In other words, it doesn't matter what you are measuring to get the numbers which are the data in your analysis; the resulting statistical significance won't change if you switch units of measure, or even if you change the thing being measured—as long as you get exactly the same measurements, that is. In a sense, statistical significance comes from the numbers themselves, not the real-world things and events that the numbers represent.

Importance, however, is very dependent on the reality behind the numbers. It is also dependent on the context of the numbers, and, indeed, the context of the analysis itself. This is not to say that "importance" is completely subjective, but rather that there is no set formula to calculate it.

So, importance is dependent on what the numbers represent, and significance is not. How are they related to each other then? Importance is dependent on statistical significance. Importance depends on statistical significance because if the numbers themselves fail to meet the criteria to qualify as statistically significant, then it doesn't really matter what they represent. They're not confirming anything, so how can we say whether they are important or not?

Significance doesn't necessarily depend on importance because it doesn't depend on the what the numbers represent. Calculations of statistical significance are completely uniformed by what it is being measured and analyzed.

But here's the twist: while the calculation of statistical significance is independent of the importance of the data, the final result—whether or not a correlation is "statistically significant"—is dependent on the importance, because the importance determines what the cut-off point for "significance" is.

While the 95% confidence level, a.k.a., "p<.05", has become a *de facto* standard in the management literature, it's a largely arbitrary one. We could make the cut-off "p<.01", "p<.001", or "p<.10", or "p<.333" for that matter. In other words, we could report with whatever level of confidence we think is appropriate, be it 99%, 99.9%, 90% or even 66.7%. [**CONFIDENCE-LEVEL CITES**]

As Cohen explains [**CITE**], for potentially important results it often makes sense to lower the criteria as to what qualifies as statistically significant. However, the inverse argument is rarely put forward—that the criteria for statistical significance should be raised for results that are likely to be trivial.

In practice, of course, neither approach is used. Criteria for statistical significance are not set based on the potential importance, or the possible triviality, of the effects being measured. Instead, the management literature adheres to the 95% confidence level, which is something of a historical left-over, despite the potential dilution of meaningful results.

> When the principles of statistical inference were established, during the early decades of the 20th century, science was a far smaller scale enterprise than it is today. In the days when perhaps only a few hundred statistical hypotheses were being tested each year, and when calculations had to be done laboriously with mechanical hand calculators, it seemed reasonable that a 5% false positive rate would screen out most of the random errors. With many thousands of journals publishing a myriad hypothesis tests each year and the ease of use of statistical software it is likely that the proportion of tested hypotheses that are meaningful (in the sense that the effect is large enough to be of interest) has decreased… (Sterne, Cox, and Smith, 2001)

In the end, the core reason that statistical significance is not strongly related to importance is because it is not required to be.

The loophole in statistical significance

The discrepancy between statistical significance and actual importance is most apparent when there are large data sets being used. Because, as Kish (1959) explained, "if the sample is large enough the most insignificant relationships will appear 'statistically significant.' "

Statistical significance can be calculated using a number of methodologies, but to help explain the strong influence of sample size on the calculations, we'll look at Student's T-test. For Student's, the value of 't' is calculated by:

```
    r*SQRT(n-2)     r = coefficient of correlation
t = -----------     n = number of data points in sample
    SQRT(1-r^2)     degrees of freedom = ( n - 2 )
```

A result is statistically significant if the value of 't' is large enough to exceed the value in a standard table, based on the desired error-level and the degrees of freedom in the regression.

In that equation, as 'n' increases 't' will increase. With a large enough sample, you will get a sufficiently large value for 't' for almost any value of 'r'. In other words, no matter how weak a correlation is, with a large enough sample, the correlation becomes statistically significant.

For example, consider two sets of data that contain correlations that are identical in magnitude—for Student's T-test, that would mean the value of 'r', the coefficient of correlation, is the same for both. The correlation from the larger data set will have the higher 't' value, so, while neither correlation is necessarily more important than the other, the result from the larger data set is more likely to be statistically significant then the result from the smaller data set.

And while seemingly arbitrary disparate outcomes can always happen at the margin, this bias in favor of larger samples is not a marginal effect. For illustration, consider that in the papers surveyed [by Locket **CITE**], sample size ranged from under 50 to over 1500. Since the ratio of two 't' values is effectively equal to the ratio of the sample sizes, the 't' value for a correlation in a study with 1500 data points would be 30 times larger than the 't' value for an equivalent correlation in a study with 50 data points. In other words, the size of an effect in the former study could be 30 times smaller than a size of an effect in the latter, but be of equal "statistical significance."

In an editorial in the *Academy of Management Journal*, Combs proposed that the increasing availability of large data sets coupled with the ease of performing null-hypothesis significance testing afforded by modern computers and software would combine to:

> …mask other shortcomings of our research designs and leave us with itty-bitty effect sizes that limit the relevance of our research. …As management scholars, can we really suggest that managers should change their decision calculus on the basis of knowledge that some new variable explains .0025 percent of the variance in organizational performance? (Combs 2010)

Combs went beyond merely speculating about a shift toward larger data sets with smaller effects, and did some comparisons between studies published in the 1987-89 volumes and in the 2007-8 volumes of the *Academy of Management Journal*. During that 20-year period, the mean sample size rose from 300 to 3,423 (excluding the three largest samples from the later period that, if included, would raise the latter mean to a whopping 7,578). At the same time, the mean correlations fell from .22, to .17 (Combs 2010). So, as expected, the ever larger sample sizes made ever smaller correlations statistically significant.

How significance obscures importance

As an added difficulty to finding research that is relevant for real-world decisions by strategic managers, the shift toward larger and larger data sets in research quietly introduces a bias against important results—which are the results in which strategic mangers are interested.

While correlations of large magnitude will be statistically significant when studied through both large and small data sets, correlations of small magnitude will only be significant when studied through large data sets. So, while a move to larger data sets doesn't increase the number of correlations of large magnitude that are published, it does increase the number of correlations of small magnitude that are published. Therefore, the ratio of large magnitude correlations to small ones will decrease. Since the importance of a correlation is often indicated by its magnitude, and is not really tied to its statistical significance, the ratio of important to trivial correlations will also decrease.

As the literature becomes saturated with more and more empirical studies of large data sets, it becomes more and more valuable to be able to be able to separate the important results from the general pool of published results.

How significance allows incorrect "conventional wisdom" to persist

If it is commonly believed that two factors are correlated, but in fact they are not, will you find articles in the management literature demonstrating that? Probably not.

The difficulty in de-bunking management myths is that the direct approach would be to run an appropriate statistical analysis and show that the two factors are no correlated. But a lack of correlation between the two is in fact, the null hypothesis, which is what tests of statistical significance compare a correlation *against*. The probability output of a statistical significance test is, in fact, the probability that the correlation is *not* zero, so how can you find the statistical significance of a result that is actually zero?

While there are ways around this problem, but to apply them a researcher has to have set out to demonstrate a lack of correlation. What if a research does not expect that the correlation is zero? Then the result will get discarded as being erroneous, especially if the statistical power of the test was not consider as part of the design of the experiment, or if it was not considered when evaluating the results.

The statistical power is particularly relevant in interpreting a lack of statistically significant results because it describes the probability of missing a correlation. E.g., if an analysis has a statistical power of 90%, then there is a 10% chance that it will fail to detect an existant correlation. But, if an experiment has failed to find statistically significant results, the implications can be turned around the other way: a statistical power of 90% means that there's a 90% chance that the expected correlation *does not actually exist*.

If the hypothesis that is well supported by theory is being tested in an empirical study, the analysis of which has a high statistical power, and the results are not statistically significant, that is an important result. Assuming that there were no flaws in the design or conduct of the the study, or in the analysis of the study, the lack of results indicates a *flaw in the theory behind the hypothesis*.

However, if you look in the management literature, what you will not find are many articles reporting a lack of statistically significant results. It doesn't matter if that's because researchers don't submit them for publication, or because they are rejected; either way, it's another sort of loophole, one that gives a pass to conventional wisdom that's completely wrong.

How to read empirical studies in management literature

What questions should a manager asks while reading a research article?

If you are a strategic manager, what questions should you be asking as you read an academic research article that reports statistical results from a regression analysis? A good place to start is with the questions that Lockett, *et al.*, [CITATION] asked as part of their survey of papers [WHERE? WHEN?]. They asked [WHAT? {reporting statistical power}, {implying significant means important}, {reporting size of effect}, {interpreting results} {???}]

In the following sections we'll look in detail at those and other questions, to explain how they are useful, when they apply, and what you can infer about an article based on the answers to them.

Ask: "Are the authors using the word 'significant' to mean 'important', and is that use justified?"

This question can be tricky because while "significant" typically does mean "important" when it has been uncoupled from "statistically", in research articles the the modifier "statistically" may still be conceptually attached to "significant" even if the literal word is not present. So, whenever the term "significant" is used, you should pause and ask yourself, "Are the authors talking about the statistics, or the effect? Are they making a statement about the existence of a correlation, or about the relative size of its effect on outcomes?" In other words, have the authors conceptually decoupled it from "statistically", or is the word still implicitly there, modifying their use of "significant"?

In their survey, Lockett, *et al.*, [**CITATION**], only 5% of the studies surveyed correctly interpreted the magnitude of effect size of the identified correlations. In the other 95% of papers, while the value of 'r' for each correlation was reported, it was not interpreted as to the importance of that correlation. All the correlations are either implicitly of equal importance, or the reader is left to erroneously infer their importance from the statistical significance.

Ask: "What was the statistical power of the analysis?"

Statistical power is related to the probability of a Type II error—that is, to the chance of failing to see a correlation that does exist. However, what makes statistical power a good way to evaluate the overall utility of an empirical study is not the actual value reported for statistical power, but rather whether or not it was reported at all.

Unlike statistical significance, calculating the statistical power of an analysis often requires interpretation of, or at least a judgement about, the nature of the effects that the study is examining. This is because one of the variables you need to calculate statistical power is the minimum correlation that you are interested in.

There are two ways to decide what size of correlation is too small, and what size is big enough, to be included in the results. One is to base the minimum on an estimation of what those correlations mean in terms of outcomes. The other is to base it on the size of effects that previous research has reported. But, of course, that second method doesn't eliminate the interpretation of, or judgement about, the nature of the effects involved—it merely shifts the task onto the researchers behind the previously published research.

So, if the statistical power is reported someone, somewhere, at some point, has indeed considered the potential importance of the results of the analysis, and has drawn a line between the potentially important and the probably trivial sizes of effect.

Lockett, *et al.*, [CITATION], found in their survey that only [PERCENTAGE] of papers reported statistical power. While the authors of the papers that didn't report statistical power probably did consider the relative importance of the results they were reporting, we can't be as confident about them doing that as we can about the the authors who did report statistical power.

Ask: "Have the authors interpreted the statistics in a meaningful way?"

One of the reasons that regression analysis is so ubiquitous is that the quantitative results of the regression, the correlations, are unitless values. They describe ratios, not measurements, so they don't need the descriptive units that measurements require, like meters, or dollars, or preference ratings, etc. This lack of units makes is easier to compare results from two experiments, because as long as the experiments measured the same fundamental construct, it doesn't matter how it was measured. In other words, if two experiments both measured, say, length as one of the variables, it doesn't matter if one reported "meters" and the other reported "hands". The results of the respective regressions will be unitless, and therefore you could directly compare them.

The regression coefficients can be reported either with or without units. "Standardized" coefficients have had their units removed, by converting all the measurements used in the analysis into something called a "z-score". This doesn't change the results of the regression, and it makes it simple to compare the influence of factors in the final regression, because you can directly compare their respective regression coefficients. The larger the standardized regression coefficient, the larger the effect on correlation.

Unstandardized correlation coefficients are expressed in terms of their respective units of measure. They cannot be compared to each other, since their units may not be comparable—how can you, for example, quantitatively compare the percent ownership by upper management of a firm and the number of patents that firm filed in a year?—but they are easier interpret into guidance for practical application.

If you wanted to use the regression from a research study as a basis for calculating predictive estimates as part of a decision making process, the most robust approach would be to use a linear model that assigns unit weights to standardized predictor variables (Wainer 1976; Cohen 1990). To do that you would need the signs of the correlation coefficients, and the descriptive statistics of all the variables.

So, if the tables of results in a study use standardized, unitless values, unless the authors have included an interpretation of those results that explains the appropriate units of measure, it is going to be difficult if not impossible to apply those results quantitatively for making management decisions.

You can qualitatively apply them, but that's rarely of much practical value—remember the hypothetical example given at the beginning? For management decisions, the existence and direction of correlations are often obvious. What's unknown is the magnitude and/or the relative importance of the correlations.

Ask: "What is the size of variation in the outcome?"

The inputs can never explain more than 100-percent of the variation in the outcome. If the complete range of variation in the outcome doesn't comprise an important amount of change, than you can't directly report that the correlation is an important one.

It is possible to extrapolate, for example, a very strong correlation from a small to a large range of variation in outcome, and make the case that the correlation is therefore an important one. But researchers should be clear about when they are doing this, and should address some of the pitfalls in doing so, e.g., they should discuss how the error in the result increases the farther it is extrapolated from the data.

Finally, a meta-question: "Why is there no research that covers what I'm looking for?"

A lack of research into the relationship between any set of factors may simply be because management researchers don't know that it's a question of interest to managers. Or there may be a lack ways to gather the necessary data.

Or, it may be because there is no statistically significant relationships between them. The unfortunate reality is that the empirical studies in the management literature are almost exclusively about correlations that are statistically significant. The statistical significance of a zero correlation can't be directly determined in an experiment that was expecting to find a non-zero correlation.

So, if there is no research that explains the relationship between a set of factors, that may mean that no relationship exists.

How to report empirical studies in management literature

A research article should include the answers to the questions that managers will ask

If you are a management researcher reporting the results of a regression analysis, what information should you include if you want your article to be relevant to managers? You should, at a minimum, include the information a manager would need to answer the questions outlined in the previous section. So, the recommendations here are some of the ways you might answer the recommend questions in that section.

For the question: "Are the authors using the word 'significant' to mean 'important', and is that use justified?"

According to Kish (1959), " 'significance' should stand for meaning and refer to substantive matter," but that ship has sailed. It is more effective to restrict the use of "significance" to refer only to statistical characteristics, than try to distinguish between "substantive significance" and "statistical significance".

Therefore, the best practice is to **never** use the word "significant" or "significance" without attaching the word "statistically" or "statistically" to it. If you are not, in fact, referring to a statistical property, use some other word. When you're revising a paper, highlight every use of "significant" or "significance" and the associated use of "statistically" or "statistically". If there isn't one, either add it, or change "significant" to something else. That may sound like a ridiculous, elementary-school-like approach, but if you want to avoid ambiguity in the minds of managers who read the paper, it is important to avoid any possible confusion between statistical significance and practical importance.

Also, when you use "important" you should be able to justify that use, with something other than a estimate of error. A lack of error does not bestow importance. You needn't couch every implication or use of "important" in a justification, but the importance of a result should be explicitly explained somewhere in the paper.

For the question: "What was the statistical power of the analysis?"

If you designed your statistical experiment from the ground up, you probably already know what the statistical power was, because that was something you used to do the design. Just be sure to report it, and to report it as having been part of the initial design.

If you didn't target a specific statistical power in the design of your statistical analysis, then it will likely be easier to address this question after you have dealt with the following one in which you will interpret the results of the analysis in practical terms. You can, once you understand the results in practical terms, establish a cut-off for the size of effect that would be worth finding, and use that as the effect size—along with the confidence level, and the sample size—to calculate the statistical power.

Avoid simply using the smallest correlation you found, or the size of effect that you expected to find. The latter is typically used when designing an experiment, the former is often used when a researcher can't be bothered to determine if the result is or isn't trivial.

For the question: "Have the authors interpreted the statistics in a meaningful way?"

Remember that this question will be asked by a practitioner, by an actual manager of a firm. So, a meaningful interpretation is this context is one that would help a manager either to make a decision, or to understand what the manager observes happening.

The first step in this is to convert the regression results into terms that a manger can understand and use, typically in one of two ways. The first way is quantitative, where you translate the regression analysis results from unitless values into values with units with which a manager would be accustomed e.g., US dollars. The second way is qualitative, where the results are explained in a binary way, e.g., "increasing" vs. "decreasing", "success" vs. "failure", [[ETCETCETC]]

From there you can establish some context for the results, explain the size of effect that your regression results represent in terms of strategic management, and justify the cut-off for importance that you applied to determine the statistical power of your analysis, or that you used to design your experiment.

For the question: "What is the size of variation in the outcome?"

If you interpret the result in a qualitative way, this question isn't really applicable. But if your meaningful interpretation of the results was quantitative, then considering the range of outcome variation becomes important, yet easy to overlook. If you normalize your data, it becomes especially easy to forget to consider the range of outcomes in your data in regard to the range of all possible outcomes, because you are artificially setting that entire range of possible variation essentially equal to the variation in your data.

Part of the process of meaningful interpretations will remove that artifice, denormalizing the data. Then you can put the outcomes back into context, and discuss the range of outcomes in the data in relation to the outcomes that a manager might encounter in actual practice.

Finally, the converse meta-question: "What if I didn't find any statistically significant results?"

When you're confident you've correctly gathered and analyzed the data in an empirical study, and yet you didn't find the correlation you were expecting to find, the first thing to consider is the statistical power of the experiment or analysis.

If the statistical power was low, then the problem may be simply that random error has obscured the correlation. Which typically means you need to redesign and/or repeat the study, increasing the statistical power as much as possible.

If the statistical power was high, the lack of a correlation may be an accurate and therefore interesting result. Can you change your hypothesis in a way that will contradict, instead of support, the theory behind it? If nothing else, you can use the statistical power to argue that the lack of results are worth publishing.

In addition to considering the statistical power, you should consider the importance of the effect size of the result in conjunction with the confidence level at which the result is statistically significant. E.g., if the result shows an important size of effect—it represent a firm's profits doubling in three years—and it is statistically significant at a confidence level of 89%, then you can argue that the importance outweighs the lower statistical significance.

In the end, the statistical characteristics should support the interpreted importance of results; they should not trump them.

< Insert Table 1 here >