Statistical Power Analysis with Microsoft Excel: Normal Tests for One or Two Means as a Prelude to Using Non-Central Distributions to Calculate Power

António Teixeira, Álvaro Rosa and Teresa Calapez
IBS – ISCTE Business School (Lisbon)

Abstract

This article presents statistical power analysis (SPA) based on the normal distribution using Excel, adopting textbook and SPA approaches. The objective is to present the latter in a comparative way within a framework that is familiar to textbook level readers, as a first step to understand SPA with other distributions. The analysis focuses on the case of the equality of the means of two populations with equal variances for independent samples with the same size.

This is the situation adopted as case 0 by Cohen (1988), a pioneer in the subject, to develop his set of tables and so, the present article can be seen as an introduction to Cohen's methodology applied to tests based on samples from normal populations. We also discuss how to extend the calculation to cases with other characteristics (cases 1 to 4), similarly to what Cohen proposes, as well as a brief discussion about the advantages and shortcomings of Excel. We teach mainly in the area of business and economics, which determines the scope of our analysis.

1. Introduction

If you browse a sample or even the universe of textbooks of statistics applied to business and economics, you will most surely see references to the calculation of the power of statistical significance tests only when the normal distribution is used (Berenson and Levine, 1996; Bernstein and Bernstein, 1999; Freund, 1962; Kazmier and Poll, 1979; Levin and Rubin, 1998; Levine *et al.*, 1999; Pestana and Velosa, 2002; Pinto and Curto, 1999; Reis *et al.*, 2001; Sanders, 1990; Sandy, 1990; Smithson, 2000; and Spiegel, 1972). Watson *et al.*, 1993 present the subject as optional, while Webster (1998) and Berenson and Levine (2004) only explain the concept without giving any examples. The calculation of the minimum sample size that guarantees predefined values for alpha and beta is a main subject in only three of the above references (Berenson and Levine, 1996, Kazmier *et al.*,

1979 and Smithson, 2000), while a fourth (Freund, 1962) presents the formula, asking for its deduction as an exercise. In brief terms, SPA does not get the importance it deserves in these books.

The cause of this situation is most surely the fact that the extension of SPA involves an additional set of (non-central) distributions and a greater number of tables. The panorama is not exactly the same in the social sciences where the greater importance given to the subject can be explained by the American Psychological Association (APA) requirements (Kline, 2004; American Psychological Association, 1994 and 2001).

As Smithson (2002)
alleges when talking about a similar subject (confidence intervals with non-central distributions), "*prior to the wide availability of computing power, exact confidence intervals for the proportion for small samples were painstakingly hand-calculated and tabulated. Confidence intervals whose underlying sampling distributions are not symmetrical and not a constant shape were labor intensive and impractical without computers, so they were (and still are) neglected in textbooks and, ironically, popular statistical computer packages.*" For example, in what concerns SPA, SPSS Inc offers a product that is independent of the SPSS application itself, Sample Size.

We intend to fill this gap by presenting an integrated analysis with a common language that will allow the development of a set of tools to be used knowingly by students and practitioners, giving them a path to think about the problem and to analyze results produced by others, eliminating the pure automation involved in obtaining numbers by pressing a key in the computer keyboard, or avoiding what Pestana and Velosa (2002) call turning off their heads when turning on the computer. With the advent of computers and the generalization of its use, computation limitations are something from the past. Excel has the capacity to implement these models. The necessary capabilities that are not built in can be created through user defined routines. It is not certain that exact results are always obtained but Excel is a privileged tool for students to perform and follow the several steps involved in the situations under consideration.

When designing a course we must be aware that students can have in their future professional lives two levels of contact with the subjects taught. One refers to performing statistical analyses and the other reading those made by others. The teacher must not consider courses with closed borders. Doors must be opened, allowing insights into things beyond the course limits. As we teach at graduation and post-graduation levels, we have been confronted with many cases of students declaring "now I understand what you really meant when saying … ." If no such doors exist, crossing to wider perspectives is more difficult.

Understanding Cohen's framework leads to understanding the effect size approach widely used in the social sciences. Even if students will never use this methodology, they may be exposed to it when reading analyses performed by others. Additionally, they may be exposed to it, especially in the fields of the social sciences, in courses like research methods. We consider that teachers who have such students in their classes might find it helpful for thinking about how to present power analysis in class.

Although table based teaching introduces limitations to the domain to be covered and in some cases returns discrete information about a continuous underlying distribution, we decided to conduct this first approach in parallel with the tables created by Cohen (1988). In this way, the present article is an introduction to Cohen's methodology applied to tests based on samples from normal populations. It could be seen as "the chapter that Cohen (1988)
did not write" or a preamble to a better understanding of his work. Our methodology will then show how one can easily construct Cohen-like tables using Excel. Besides constituting a tribute to Cohen's pioneer work, we also believe this is the best way to convey the knowledge at the level we want at this stage, which is to understand the methods involved and the tools available instead of supplying automated solutions subject to the "garbage in garbage out" perils. We do not intend to follow the instant pudding route. Even if the user has access to software applications that will directly give the desired results, we intend to arm him with enough knowledge and alternative routes to check those results by performing the calculation process step by step.

Tables can also be seen as a base for charting results. In this specific case, there is a wide range of combinations of inputs and formulas to use. When comparing functions and routines used, these tables define what we can call a *Cohen's universe*, fixing borders. We can confine accuracy analyses to this universe, especially when dealing with the subject for teaching purposes.

Furthermore, computing power directly does not replace the use of tables in teaching. Instead of static tables we can have applications that will reformulate the tabulated values each time entry values are changed. For example, in SPA, the teacher can generate problem-specific tables and include them in examination papers instead of generic tables that would require a large variety of entries. This kind of models and reasoning tends to assume a more important role, as classes - and even examinations – are becoming more and more lab based. The construction of tables is also a good exercise for students to become familiar with the characteristics of statistical distributions.

As this article is Excel based, we also include a section about software limitations, somewhat beyond the scope of this specific article, in order to create a centralized reference to this matter to be used in forthcoming work. This article is just the first of a series about power and precision analysis. The concentration of these considerations helps to confer unity to the set of final articles.

To avoid unnecessary extension of the article, we assume that the reader is familiar with concepts such as type I (alpha) and type II (beta) errors, structure of hypotheses testing and the power (pi) of a test, avoiding the need to explain them in detail. The meaning of all symbols used can be found at the end of the article. This results in a more compact and easier to read article.

2. Excel Advantages, Shortcomings, and Oddities

The wide use of Excel is the main reason for its choice. It is a tool familiar to most users for the development of models related to a wide range of scientific and practical fields. Despite the limitations we are identifying here, it is an excellent tool for teaching purposes, allowing students to shed light into what can otherwise be seen as black boxes. With Excel one can learn while performing all calculation steps and easily creating graphical representations. On the other side, the shortcomings we are listing emphasize the necessity of analyzing critically computer generated results even in applications that are considered as more reliable.

There are two eras concerning the use of Excel for statistical analysis: pre and post Excel 2003. In his review of Excel 2003, Walkenbach (2008), referring to what Microsoft indicates as enhancements in the statistical functions, states: "*This, of course, is a new definition of the word enhancements. Truth is, these functions have been broken for more than a decade, and they finally return accurate results in Excel 2003.*"

Knüsel (1998, 2002, 2005) has been studying the accuracy of Excel statistical functions and considers that "*some of previously indicated errors in Microsoft Excel 97 and Excel XP have been eliminated in Excel 2003. But some others have not been corrected … and new ones have been found …*". Sleeper (2006) approaches this subject in a subchapter named *Defects and Limitations of Microsoft Excel Spreadsheet Software* where he presents a list of problems either solved or remaining in Excel 2003, indicating that he still found the latter in the beta version of Excel 2007. We have checked the cases indicated in Knüsel (2005) and found that nothing has changed in the release version.

In statistical terms the remaining problems concern the accuracy of statistical functions in part of their domain, namely, using Excel designations, in RND (in what concerns the random number generator), POISSON, BINOMDIST, CHIINV, GAMMADIST, BETAINV, and NEGBINOMDIST.

Sleeper (2006)
also refers to the ill-functioning of Excel's QUARTILE and PERCENTILE functions. It is not quite so. As

Larson and Faber (2000)
acknowledge, there are several ways of finding the quantiles of a data set. Instead of the most usual formulas to calculate the position of the first and third quartiles, $(n+1)/4$ and $3(n+1)/4$, Excel uses $(n+3)/4$ and $(3n+1)/4$ respectively. These seldom appear in textbooks but can be found, for instance, in Murteira *et al*. (2002). The method should be included in the help file to avoid misinterpretations.

Another shortcoming we did not find reported is the lack of ability to detect multimodal data: the MODE function, given any data set, returns only one value. The problem is that, for multimodal data, the value reported may vary upon sorting.

As a conclusion Sleeper (2006)
acknowledges that for statistical calculations Excel 2003 and later are not perfect but acceptable, recommending that the Analysis Toolpak add-in provided with any version of Excel should never be used, giving as reference the Microsoft Knowledge Base Article to get a detailed view of the problems involved (Microsoft, 2007), taking into account that some of them have been improved.

Besides the limitations indicated above there is a wider set of oddities, some of them presented by Walkenbach (2008)
regarding the sum of Boolean values. We also found that validation of input does not work when using the *Paste Special* alternative. Microsoft (2007a, 2007b) elaborates about problems involving floating-point arithmetic and how to remedy them.

Whenever a new application is developed bugs tend to appear. Nevertheless, Excel is a valuable and versatile tool to accomplish a wide range of objectives including the one we are pursuing: to teach statistics.

3. Statistical Power Analysis with the Normal Distribution Using Microsoft Excel

There are several different situations involving the tests for the mean of one or two populations. Cohen (1988) considers those indicated in Table 1.

Table 1 - Tests for means included in Cohen (1988)

Case 0	2 means; $\sigma_a = \sigma_b$; $n_a = n_b$
Case 1	2 means; $\sigma_a = \sigma_b$; $n_a \neq n_b$
Case 2	2 means; $\sigma_a \neq \sigma_b$; $n_a = n_b$
Case 3	1 sample size n
Case 4	2 paired observations in a sample size n

He develops the tables for case 0, although they can also be used for the other cases provided some additional calculations are made to find the tables' input values. Consequently, we will deal with this case as the base situation. The tests can also be one tailed (upper or lower) or two tailed. Cohen presents tables for two and one upper tailed tests (obviously, the latter will allow the calculation of power of one lower tailed test). All the other cases can be transformed into one of these two. It is notable that Cohen leaves aside the case in which $\sigma_a \neq \sigma_b$; $n_a \neq n_b$. We will also ignore it for now, leaving its consideration for future work on power and precision while testing means.

Due to its similarity, we will deal only with two tailed tests, using the upper tail to calculate the power. The case we consider is then the one on the left in Table 2, where H_p and μ_p refer to the values of the mean to use when calculating power. This restriction does not affect the generality of the conclusions.

<div align="center">

Table 2 - Type of tests considered by Cohen

Two Tailed	Upper Tailed
$\begin{cases} H_0 : \mu_a - \mu_b = \mu_0 \\ H_1 : \mu_a - \mu_b \neq \mu_0 \end{cases}$ $H_p = \mu_a - \mu_b = \mu_p$ $\alpha = \{0.01 ; 0.05 ; 0.10\}$	$\begin{cases} H_0 : \mu_a - \mu_b = \mu_0 \\ H_1 : \mu_a - \mu_b > \mu_0 \end{cases}$ $H_p = \mu_a - \mu_b = \mu_p$ $\alpha = \{0.01 ; 0.05 ; 0.10\}$

</div>

The test statistic can be either the sample mean $\overline{X}$ or its standardized value $z_{\overline{X}}$ in the case of one population whereas in the case of two populations, it can be the difference between the two sample means $\overline{X}_a - \overline{X}_b$ or its

standardized value $z_{\overline{X}_a - \overline{X}_b}$
. Textbooks show both, standardized and non-standardized, to arrive at a decision, but the value of the type II error, beta, is calculated (almost always) using the non-standardized test statistic. It happens that Cohen's methodology asks for the opposite, that is, the use of the standardized value.

In any case, the test is carried out by defining a rejection (R) and a non-rejection $(\overline{R})$ interval that are expressed in terms of the statistic used for the test. As we will calculate power as $\pi = 1 - \beta$, we are concerned with $\overline{R}$, which is defined in Table 3. Note that we are using $]a;b[$ for an open interval, $[a;b]$ for a closed one, and every mixed situation accordingly with these.

<div align="center">

Table 3 – Rejection and non-rejection regions

Test Statistic	Two Tailed	Upper Tailed
Unstandardized	$\overline{R} = \left] \left(\overline{X}_a - \overline{X}_b\right)_{lcv} ; \left(\overline{X}_a - \overline{X}_b\right)_{ucv} \right[$	$\overline{R} = \left] -\infty; \left(\overline{X}_a - \overline{X}_b\right)_{ucv} \right[$
Standardized	$\overline{R} = \left] - z_{lcv} ; z_{ucv} \right[$	$\overline{R} = \left] -\infty ; z_{ucv} \right[$

</div>

We can approach power from two perspectives: (1) *a posteriori*: calculating the power after performing the test; (2) *a priori*: calculating the sample size needed to obtain a minimum desired value for power, as well as for alpha. Additionally, we will consider only the cases in which the finite population correction factor for the calculation of the standard error is not necessary.

3.1 Computation of Power Using the Usual Way

As an example of calculating power for the two tailed test alternative, we will use the specific situation presented in Table 4.

Table 4 – Example considered in the analysis of power calculation

An investor is interested in testing the hypothesis that the mean household monthly income is the same in two metropolitan areas. For this purpose, a statistical consultant performed a survey using two independent simple

random samples with the same size: $n = n_a = n_b = 36$, obtaining the following results for the mean of each sample:

$\overline{X}_a = \$1576.962 \quad \overline{X}_b = \$1410.144$.

From the results of previous studies, the consultant assumes that the standard deviation of the household incomes in both areas is the same:

$\sigma = \$600$ The value $\alpha = 0.05$ is used for the statistical significance of the test. The alternative value of the difference between the two means for

computing the power of the test is $\mu_y = 450$.

3.1.1. Power Computation *a posteriori*

The *usual way* of performing this test using $\overline{X}_a - \overline{X}_b$ as the test statistic leads to the results presented in Table 5. Putting aside for now the calculation of beta, the decision is taken based on the assumption that H_0 is true, i.e. there is no difference between the two population means. From Table 5 we can see that the test statistic belongs to the non-rejection interval, leading to the non-rejection of H_0.

Table 5 – Performing the test using $\overline{X}_a - \overline{X}_b$ as the test statistic

$\begin{cases} H_0 : \mu_a - \mu_b = 0 \\ H_1 : \mu_a - \mu_b \neq 0 \end{cases}$ $H_y = \mu_a - \mu_b = 450$ $\alpha = 0.05$	$\sigma_{\overline{X}_a} = \sigma_{\overline{X}_b} = \dfrac{\sigma}{\sqrt{n}} = 100$ $\sigma_{\overline{X}_a - \overline{X}_b} = \sqrt{\sigma^2_{\overline{X}_a} + \sigma^2_{\overline{X}_b}} = 141.42$
$\left(\overline{X}_a - \overline{X}_b\right) = 166.818$ $\left(\overline{X}_a - \overline{X}_b\right)_{lcv} = \left(\mu_a - \mu_b\right)_{H_0} - z_{1-\alpha/2}\, \sigma_{\overline{X}_a - \overline{X}_b} = -277.181$ $\left(\overline{X}_a - \overline{X}_b\right)_{ucv} = \left(\mu_a - \mu_b\right)_{H_0} + z_{1-\alpha/2}\, \sigma_{\overline{X}_a - \overline{X}_b} = +277.181$	
Decision: do not reject H$_0$	

The probability of committing a type II error (beta) is associated with the non-rejection interval of the normal curve represented at the left on Figure 1. However, when computing the type II error, the falsity of H_0 is assumed, meaning that we can no longer work with a normal distribution with $\mu_0 = (\mu_a - \mu_b)_{H_0} = 0$.

Therefore, the calculation of power needs specification of alternative values for the difference between the two means. The situation can be dealt (as we do in this case) either with the consideration of a point value indicated in H_p, or by constructing a power function for several alternative values for the mean. We followed the former because the later is just a repetition of the procedure presented.

Type I error (alpha) is represented by the darker (blue) shaded area in Figure 1 and Type II error (beta) by the lighter (grey) shaded area, with value

$$\beta = P\left[\left(\overline{X}_a - \overline{X}_b\right)_{lcv} < \left(\overline{X}_a - \overline{X}_b\right) < \left(\overline{X}_a - \overline{X}_b\right)_{ucv} \mid \mu = \mu_p; \sigma = \sigma_{\overline{X}a-\overline{X}b}\right] = 0.11$$

The power of the test, $\pi = 1 - \beta = 0.89$, is represented by the brown area in Figure 1.

Figure 1 – Graphical representation of alpha, beta and power
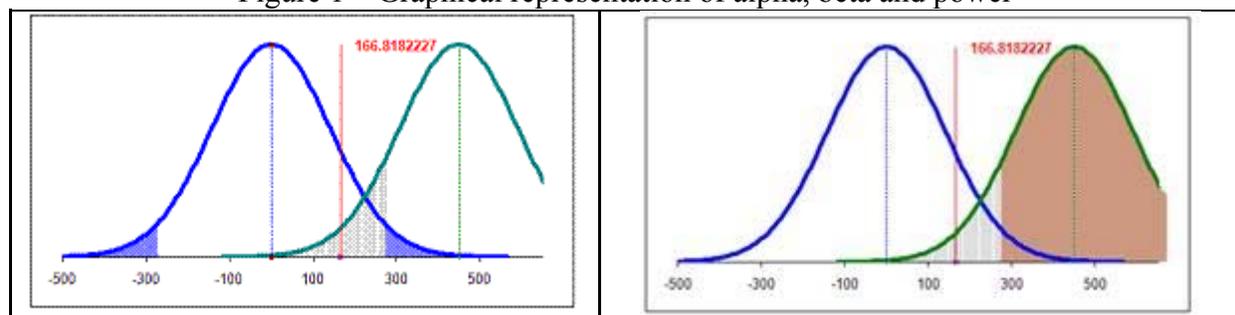
Figure 2 shows how this value can be easily obtained using the functions embedded in Excel. Please note that in the Excel sheets pictured, grey background cells correspond to inputs, white background cells to intermediate results and inverse background (yellow over blue) cells to final results.

Figure 2 – Computation *a posteriori* of power using Excel formulas and embedded functions

3.1.2. Power Computation *a priori*

The *a priori* computation of power is the determination of a sample size n guaranteeing that pre-defined values for alpha and beta will not be exceeded. Let us suppose that we want to test the equality of the population means in a way that the values of alpha and beta will not be greater than *0.05* and *0.10* respectively. Figure 3 illustrates the base from which we can derive a formula to compute the minimum *n*.

Figure 3 – Basis to derive the sample size given maximum alpha and beta values
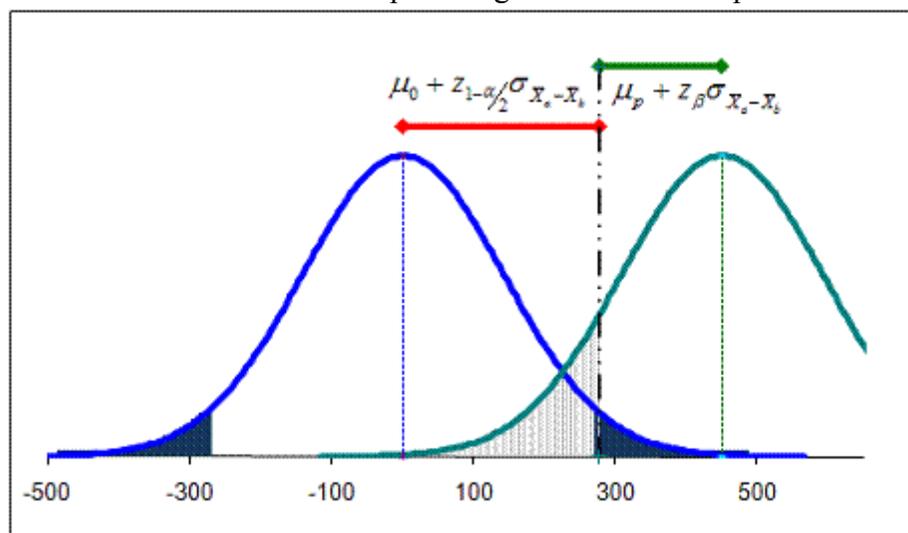
From

$$\mu_p + z_\beta \sigma_{\bar{X}_a - \bar{X}_b} = \mu_0 + z_{1-\alpha/2} \sigma_{\bar{X}_a - \bar{X}_b}$$

(1)

and

$$\sigma^2_{\bar{X}_a - \bar{X}_b} = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2\sigma^2}{n}$$

(2)

It follows that

$$n = \frac{(z_{1-\alpha/2} - z_\beta)^2 \, 2\sigma^2}{(\mu_p - \mu_0)^2}$$

(3)

As we are only considering cases in which the beta value is no greater than *0.50*, then z_β will always be negative, whereas $z_{1-a/2}$ will always have the opposite sign. This is reflected in the graphic since the upper

critical value, $\mu_0 + z_{1-\alpha/2} \, \sigma_{\bar{X}_a - \bar{X}_b}$, is always between the values of μ_0 and μ_p: we have to sum some *positive* quantity to μ_0 in order to reach it. In the same way, we have to sum some *negative* quantity to μ_p to obtain the same value. Consequently, the subtraction ($z_{1-a/2} - z_\beta$) can be seen as the sum of two values of the same sign, causing its square to be equal to the squared sum of the absolute values. Thus, the above formula (3) can be rewritten as:

$$n = \frac{\left(|z_{1-\alpha/2}| + |z_\beta|\right)^2 \, 2\sigma^2}{(\mu_p - \mu_0)^2}$$

(4)

The advantage of using (4) instead of (3) is that it still holds if the power is calculated using the lower tail, since

$|z_{1-\alpha/2}| = |z_{\alpha/2}|$ and $|z_\beta| = |z_{1-\beta}|$. Also note that (4) gives us in fact a lower bound on *n*, so we should take as minimum sample size the first integer greater or equal to the value obtained.

This formula is very easy to calculate in Excel, as it can be seen in Figure 6. For the example given, taken an

alpha of *0.05* and a beta of *0.10,* the sample size for each population's sample should be no lower than *38*.

3.2 Computation of Power Using the Standardized Value

3.2.1. Power Computation *a posteriori*

Another way of performing the test is using the standardized normal distribution, i.e., adopting $z_{\bar{X}_a - \bar{X}_b}$ as the test statistic. For the example above the intervals of rejection and non-rejection would be in this case:

$$R = \left]-\infty;-1.96\right] \cup \left[1.96 ; +\infty\right[\quad \therefore \quad \bar{R} = \left]-1.96 ; 1.96\right[\tag{5}$$

and

$$\bar{X}_a - \bar{X}_b = 166.81 \Rightarrow z_{\bar{X}_a - \bar{X}_b} = 1.18 \Rightarrow z_{\bar{X}_a - \bar{X}_b} \in \bar{R} \tag{6}$$

As the decision must be the same as before, the standardized value lies in the non-rejection region. To compute the type II error, the remarks made previously remain valid: it is a matter of calculating the probability of occurrence of the non-rejection interval with a normal curve that is not the one corresponding to the validity of

H_0. Only now we are using a different variable, $z_{\bar{X}_a - \bar{X}_b}$, which is standardized only under the validity of the null

hypothesis. As, for the computation of power, we must assume the falsity of H_0, then $z_{\bar{X}_a - \bar{X}_b}$ is no longer standardized, but still has unitary variance. For the example given, the alternative normal curve has a mean of

$$\mu = \frac{450 - 0}{141.42} = 3.18 \tag{7}$$

We will use an oxymoron to name it: *non-central standardized normal distribution*, as it will be useful in the more general framework we want to develop, although that distribution does not exist (if it is standardized, its mean is zero). It suggests that we are referring to a normal distribution in terms of *z*, but with a non-zero mean of $\mu = 0 + \delta$. Although this is not standard terminology, it will help facilitate understanding the (central) *t versus* non-central *t* distribution.

Unlike other distributions, non-central and central normal distributions have exactly the same shape, that is, non-centrality does not affect the curve, except for a horizontal translation along the *z* axis. This is why computing power is much easier when the normal distribution is concerned. In this case:

$$\beta = P\left(-1.96 < z < +1.96 \mid \mu = 3.18, \sigma = 1\right) \tag{8}$$

To completely specify a non-central standardized normal distribution it is only necessary to define the non-centrality parameter, δ. In the present case, this parameter equals the standardized difference of means (7), i.e.

$$\delta = \frac{|450 - 0|}{141.42} = 3.18 \quad \Rightarrow \quad \delta = \mu \tag{9}$$

Note that: i) the non-central distribution tends to the correspondent central distribution as δ tends to zero; and ii) in the above formula, the denominator is the standard error of the difference between the two means because what is desired is the distribution of this variable.

We can use Excel in two different ways to calculate power using *z* as the test statistic. The first one is

conceptually *pure*, adding the non-centrality parameter to the mean (zero in this case), as illustrated in Figure 4.

Figure 4 – Power computation *a posteriori*: version 1, conceptual

The other way is to *decentralize* the critical values of *z*
by an amount δ and use the central standardized normal distribution to calculate probabilities, as shown in Figure 5:

Figure 5 – Power computation *a posteriori*: version2, computational

The first option is conceptually sound while the second facilitates the introduction of formulas by using the NORMSDIST instead of the NORMDIST function.

3.2.2. Power Computation *a priori*

Equation (4) shows that the formula to compute *a priori* power does not contain the test statistic. This means that it can be applied to either $\overline{X}$ or $z_{\overline{X}}$.

3.3 Computation of Power Using the Effect Size

Instead of introducing a non-centrality parameter, δ, Cohen, when formulating his SPA framework, defined the *effect size d* in answer to the necessity he felt of "*a 'pure' number, one free of our original measurement unit, with which to index what can be alternately called the degree of departure from the null hypothesis of the alternate hypothesis, or ES (effect size) we wish to detect*". The effect size can be seen as the extent to which the two populations do not overlap (it is the extent to which the experiment has an effect of separating the two populations). In this case what is involved is the distribution of the original random variable and not a statistic obtained from sampling. Thus, the measure of dispersion used for standardization is the standard deviation of that variable ($\sigma_A = \sigma_A = \sigma$).

For the two independent samples case, with homogeneous variances, this effect size is:

$$d = \frac{|\mu_p - \mu_0|}{\sigma} \tag{10}$$

Consequently:

$$\delta = d \sqrt{\frac{n}{2}} \tag{11}$$

Cohen (1988) also presents other measures of the effect size (U_1, U_2, U_3, r and r^2). We do not discuss these here. Cohen suggests that, in the absence of information, one might chose to fix the effect size as small (d=0.2), medium (d=0.5) or large (d=0.8) depending upon the subject studied.

3.3.1. Power Computation *a posteriori*

One of the advantages of working with effect sizes and standardized statistics is that we no longer need to specify explicitly the alternative value, μ_p, to consider for the calculation of power. We only have to give, as input, alpha, the sample size n and the effect size considered.

For the example we are following, with a sample size of 36 and an alpha value of 0.05, we can either fix an alternative value of 450 (as we have done previously) or, equivalently, chose an effect size of *0.75* (large). Type

II error, and hence power, can then be easily computed in Excel, as d=*0.75* gives $\delta = \mathbf{0.75} \times \sqrt{\dfrac{\mathbf{36}}{\mathbf{2}}} = \mathbf{3.182}$ and β=0.111 as seen on Figure 5.

This relationship allows the construction of general tables, which we will refer to as Cohen-like tables, presented in section 4.

3.3.2. Power Computation *a priori*

The formula to compute the sample size that guarantees minimum pre-defined values of alpha and beta is very similar to the one used when working with the unstandardized statistic and it can be deducted directly from (4):

$$n = \frac{\left(\left|z_{1-\alpha/2}\right| + \left|z_\beta\right|\right)^2 2\sigma^2}{\left(\mu_p - \mu_0\right)^2} \qquad d = \frac{|\mu_p - \mu_0|}{\sigma} \Longleftrightarrow \qquad n = \frac{2\left(\left|z_{1-\alpha/2}\right| + \left|z_\beta\right|\right)^2}{d^2} \tag{12}$$

Figure 6 - Computational power computation *a priori* based on the effect size

4. How Cohen Deals With the Other Cases

Cohen's tables consider case 0 for *a posteriori* and *a priori* power. Let us see the methods he proposes to calculate power for the other cases based on the results of the tables, recalling that we are analyzing only two tailed tests using the upper tail when the value of beta is not greater than *0.50*.

4.1 *A Posteriori* Power

Adopting Appendix B's Cohen-Like Table as reference, Table 6 presents the formulas to get the inputs for reading *a posteriori* power for all cases. The inputs for Appendix B's Table are the effect size, *d*, and the sample(s) size *n*, while the output, a posteriori power π, is expressed as a percentage. Instead of computing *d*, one can always choose directly a magnitude for the effect size, as proposed by Cohen.

Table 6 – Applying Cohen-Like table for *a posteriori* power to all cases

INPUTS						
Case	**s**	**n**	**d**			
			Formula	**Small**	**Medium**	**Large**
0	$\sigma = \sigma_a = \sigma_b$	$n = n_a = n_b$	$d = \dfrac{\left\vert \mu_p - \mu_0 \right\vert}{\sigma}$	0.2	0.5	0.8
1	$\sigma = \sigma_a = \sigma_b$	$n = \dfrac{2 n_a n_b}{n_a + n_b}$ (harmonic mean)	$d = \dfrac{\left\vert \mu_p - \mu_0 \right\vert}{\sigma}$	0.2	0.5	0.8
2	$\sigma = \sqrt{\dfrac{\sigma_a^2 + \sigma_b^2}{2}}$	$n = n_a = n_b$	$d = \dfrac{\left\vert \mu_p - \mu_0 \right\vert}{\sigma}$	0.2	0.5	0.8
3	$\sigma = \sigma_a$	$n = n_a$	$d = \dfrac{\left\vert \mu_p - \mu_0 \right\vert}{\sigma} \sqrt{2}$	0.2	0.5	0.8
4	Case 3 methodology can be applied for a variable representing the difference between the paired observations: $$X_c = X_a - X_b$$					

4.2 *A Priori* Power

The starting point is the Cohen-Like Table presented in Appendix C. As before, we can either compute the effect size, *d*, or choose a small, medium or large magnitude of effect. After fixing the power to be obtained, π, we can directly read a value, *n*, from the Appendix C Table. Table 7 shows the relationship between this value and the sample(s) size(s) to consider.

Table 7 – Applying Cohen-Like table for *a priori* power to all cases

Case	Sample(s) size(s)
0	$n_a = n_b = n$
1	*Fix n_a (> n/2)* $n_b = \dfrac{n_a n}{2 n_a - n}$
2	$n_a = n_b = n$
3	$n_a = n$
4	Case 3 methodology can be applied for a variable representing the difference between the paired observations: $X_c = X_a - X_b$

5. Effect Size, Power and Precision in Statistical Inference

When conducting inference for population mean(s) we can follow two mutually exclusive routes: power or precision analysis. The former is done through testing hypotheses and the latter by the construction of confidence intervals. They should not be mixed. There is a mathematical similarity between the two approaches, namely in dealing with variance, but they are addressing different questions. Our goal is to deal with power analysis. This reference is just a warning. Borenstein (2008) states that it is wrong to construct confidence intervals to estimate a parameter value (precision approach) when the sample size is obtained by fixing the maximum value for the type II error (power approach) as illustrated in Figure 7. In the last case the confidence interval must be used to make a decision about the null hypothesis.

Figure 7 – Confidence intervals as a tool shared by power and precision analyses
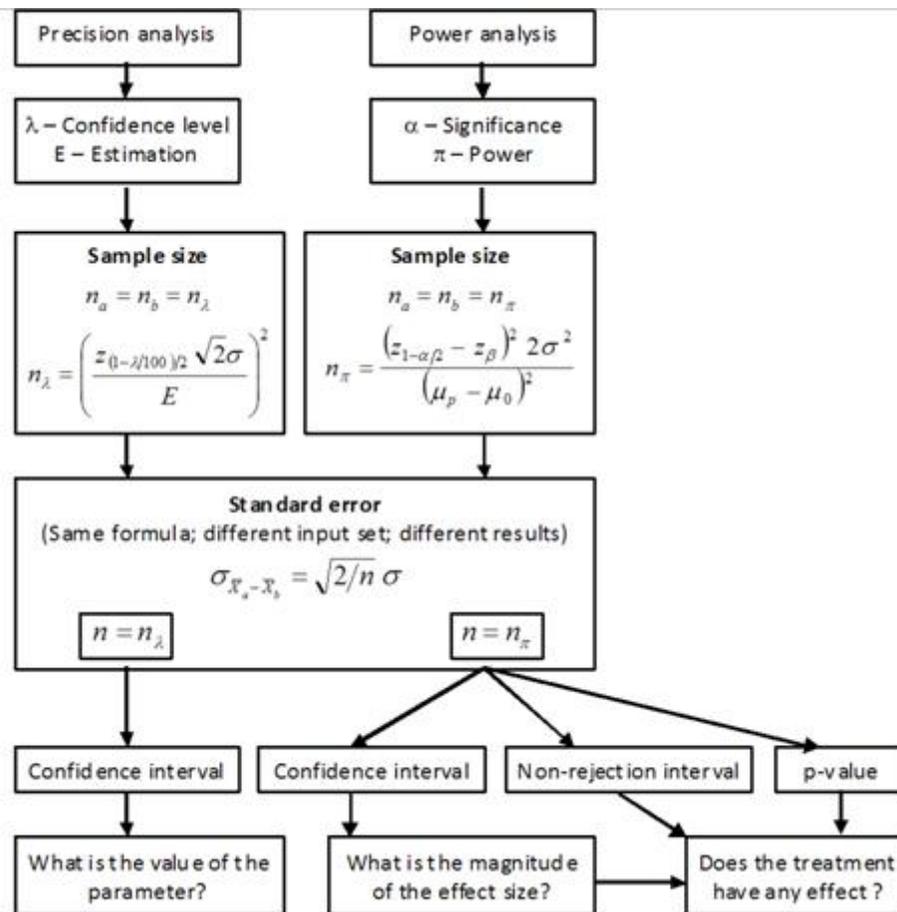
He considers that power analysis gives power for a specific effect size: "*For example, the researcher might report 'If the treatment increases the response rate by 20 percentage points the study will have power of 80% to yield a significant effect.'*" In simple terms, he says that "*the effect size used in the power analysis should represent the smallest effect that would be of clinical substantive significance*."

A chain store intends to open a new outlet in some region A, if it is possible to assume an average household income of \$50,000 for that area. In this case, the business will surely not be affected if the average household income is \$49,500. Then, how large should the difference be in order to decide that it is not worth opening the new outlet? It depends mainly on the knowledge the researcher has about the subject being studied and must be defined at the start of an SPA in terms of the effect size (a standardized index). This definition is one of the factors that affects the size of the sample in what we call the design of the sample for *a priori* power. We can also see power as our prediction of what proportion of the studies will yield a statistically significant effect (Borenstein, 2008).

A study designed to detect a large effect will not have adequate power to detect a small effect. A study designed to detect a small effect will have power enough to detect a large effect. In this last case the power is greater than needed, that is, the sample is greater than necessary.

6. The Effect Size Approach in Class

We have been introducing some changes in the structure and methods in one of our courses in order to achieve a greater cohesion in its contents, as well as to diminish the number of formulas used (that in some cases are no more than *variations on the same theme*). For example, instead of treating separately sampling distributions, confidence intervals and hypotheses testing, we deal first with the case of one mean for all those aspects, and

then we present the case for two means and independent samples, and so on. At every step the focus is on the calculation of the respective standard error.

Until now the experience has been positive, giving students an easier and smaller set of formulas to deal with. In what concerns the computation of beta, we anticipate that this framework will also simplify its computation – or at least without significantly increasing its complexity – while opening doors to new areas. Our previous experience in teaching it following the usual textbooks revealed great difficulties for students to acquire the desired competence. The result was a poor performance both in understanding the subject and in correctly answering questions during examinations.

For example, using only the expressions in (13) to compute type I and II errors (for one mean) increased markedly the quality of examination answers:

$$\alpha = P\left(\text{rejection region} \mid \mu = \mu_0 ; \sigma = \sigma_{\bar{X}}\right)$$
$$\beta = P\left(\text{non - rejection region} \mid \mu = \mu_p ; \sigma = \sigma_{\bar{X}}\right) \tag{13}$$

What used to be one of the difficult questions in examinations became a trivial one.

Now we are planning to take a further step forward, i.e., moving towards the effect size approach for the calculation of power with other distributions besides the normal. We hope it will be a simple task for our students, not only because of the unified framework we use, but also because they are already accustomed to using Excel in this context. The introduction of the effect size approach to SPA leads to the revision of the formulas presented in (13) to

$$\alpha = P\left(\text{rejection region} \mid \mu = 0; \sigma = 1\right)$$
$$\beta = P\left(\text{non - rejection region} \mid \mu = \delta; \sigma = 1\right) \tag{14}$$

7. Conclusion and Further Research

This is a first article in a series that intends to spread statistical teaching under the statistical power and precision analysis perspective. To attain this objective we intend: (1) to develop simple tools to deal with these subjects, integrated in a theoretical framework explained in a way that makes it accessible to textbook level readers, and (2) to analyze the current structure of courses in statistics for business and economics in order to suggest changes in their structures so that they accommodate better to this new perspective.

The present article is confined to statistical power analysis and aims to be a bridge between the *usual way* – which can only be used within testing using the normal distribution – and Cohen's framework for statistical power analysis. This first step does not leave familiar territory, while showing a new way to act that has the advantage of being able to deal with more general situations, namely, power with non-central distributions. That will be the object of forthcoming articles, where we will present both ready to use tools and underlying concepts.

Appendix A – List of Symbols

a　Probability of incurring in a type I error; statistical significance of the test

β　Probability of incurring in a type II error

δ　Non centrality parameter

μ_a　Population A mean

μ_b　Population B mean

μ_0　Value for the mean in H_0

μ_p　Value for the mean in H_p

π　Power of the test

σ　Standard deviation of a population

σ_a　Standard deviation of population A

σ_b　Standard deviation of population B

$\sigma_{\bar{x}_a}$　Standard error of the mean for samples taken from population A

$\sigma_{\bar{x}_b}$　Standard error of the mean for samples taken from population B

$\sigma_{\bar{x}_a - \bar{x}_b}$　Standard error of the difference between the two means

d　Effect size; Cohen's d

d_c　Sampled d values necessary for significance

H_0　Null hypothesis

H_1　Alternative hypothesis

H_p　Alternative value of the population mean to compute power

lcv　Lower critical value

ucv　Upper critical value

n　Sample size

n_a　Size of the sample taken from population A

n_b　Size of the sample taken from population B

n_λ　Size of the sample determined for precision analysis

n_π　Size of the sample determined for power analysis

R　Rejection interval

$\bar{R}$　Non rejection interval

$\overline{X}$　Sample mean

$\overline{X}_a$　Mean of the sample extracted from population A

$\overline{X}_b$　Mean of the sample extracted from population B

z　Standardized value of a variable

$z_{\overline{X}}$　Standardized value of the sample mean

$z_{\overline{X}_a-\overline{X}_b}$　Standardized value of the difference between the sample means

Appendix B – Cohen-Like Tables (1)

Computation of *a posteriori* power (values presented as percentages)

Normal distribution, two tails, $\alpha = 0.05$, $Ho : \mu_1 = \mu_2$

n	d_c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1	1.2	1.4
8	0.98	5	7	9	13	17	22	29	36	52	67	80
9	0.92	6	7	10	14	19	25	32	40	56	72	84
10	0.88	6	7	10	15	20	27	35	43	61	77	88
11	0.84	6	8	11	16	22	29	38	47	65	80	91
12	0.80	6	8	11	17	23	31	40	50	69	84	93
13	0.77	6	8	12	18	25	33	43	53	72	86	95
14	0.74	6	8	12	18	26	35	46	56	75	89	96
15	0.72	6	9	13	19	28	38	48	59	78	91	97
16	0.69	6	9	14	20	29	40	51	62	81	92	98
17	0.67	6	9	14	21	31	42	53	65	83	94	98
18	0.65	6	9	15	22	32	44	56	67	85	95	99
19	0.64	6	9	15	23	34	46	58	69	87	96	99
20	0.62	6	10	16	24	35	48	60	72	89	97	99
21	0.60	6	10	16	25	37	49	62	74	90	97	100
22	0.59	6	10	17	26	38	51	64	76	91	98	100
23	0.58	6	10	17	27	40	53	66	77	92	98	100
24	0.57	6	11	18	28	41	55	68	79	93	99	100
25	0.55	6	11	19	29	42	56	70	81	94	99	100
26	0.54	7	11	19	30	44	58	71	82	95	99	100
27	0.53	7	11	20	31	45	60	73	84	96	99	100
28	0.52	7	12	20	32	46	61	75	85	96	99	100
29	0.51	7	12	21	33	48	63	76	86	97	100	100
30	0.51	7	12	21	34	49	64	77	87	97	100	100
31	0.50	7	12	22	35	50	66	79	88	98	100	100
32	0.49	7	13	22	36	52	67	80	89	98	100	100
33	0.48	7	13	23	37	53	68	81	90	98	100	100
34	0.48	7	13	24	38	54	70	82	91	98	100	100
35	0.47	7	13	24	39	55	71	83	92	99	100	100
36	0.46	7	14	25	40	56	72	84	92	99	100	100
37	0.46	7	14	25	41	58	73	85	93	99	100	100
38	0.45	7	14	26	41	59	74	86	94	99	100	100
39	0.44	7	14	26	42	60	75	87	94	99	100	100
40	0.44	7	15	27	43	61	77	88	95	99	100	100
42	0.43	7	15	28	45	63	79	89	96	100	100	100
44	0.42	8	16	29	47	65	80	91	96	100	100	100
46	0.41	8	16	30	48	67	82	92	97	100	100	100
48	0.40	8	17	31	50	69	84	93	97	100	100	100

n	d_c	d										
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1	1.2	1.4
50	0.39	8	17	32	52	71	85	94	98	100	100	100
52	0.38	8	18	33	53	72	86	95	98	100	100	100
54	0.38	8	18	34	55	74	88	95	99	100	100	100
56	0.37	8	18	35	56	75	89	96	99	100	100	100
58	0.36	8	19	37	58	77	90	96	99	100	100	100
60	0.36	9	19	38	59	78	91	97	99	100	100	100
64	0.35	9	20	40	62	81	92	98	99	100	100	100
68	0.34	9	21	42	65	83	94	98	100	100	100	100
72	0.33	9	22	44	67	85	95	99	100	100	100	100
76	0.32	9	23	46	69	87	96	99	100	100	100	100
80	0.31	10	24	48	72	89	97	99	100	100	100	100
84	0.30	10	25	49	74	90	97	100	100	100	100	100
88	0.30	10	26	51	76	91	98	100	100	100	100	100
92	0.29	10	27	53	77	92	98	100	100	100	100	100
96	0.28	11	28	55	79	93	99	100	100	100	100	100
100	0.28	11	29	56	81	94	99	100	100	100	100	100
120	0.25	12	34	64	87	97	100	100	100	100	100	100
140	0.23	13	39	71	92	99	100	100	100	100	100	100
160	0.22	15	43	77	95	99	100	100	100	100	100	100
180	0.21	16	48	81	97	100	100	100	100	100	100	100
200	0.20	17	52	85	98	100	100	100	100	100	100	100
250	0.18	20	61	92	99	100	100	100	100	100	100	100
300	0.16	23	69	96	100	100	100	100	100	100	100	100
350	0.15	26	75	98	100	100	100	100	100	100	100	100
400	0.14	29	81	99	100	100	100	100	100	100	100	100
450	0.13	32	85	99	100	100	100	100	100	100	100	100
500	0.12	35	89	100	100	100	100	100	100	100	100	100
600	0.11	41	93	100	100	100	100	100	100	100	100	100
700	0.10	46	96	100	100	100	100	100	100	100	100	100
800	0.10	52	98	100	100	100	100	100	100	100	100	100
900	0.09	56	99	100	100	100	100	100	100	100	100	100
1000	0.09	61	99	100	100	100	100	100	100	100	100	100

Appendix C – Cohen-Like Tables (2)

Computation of *a priori* power

Normal distribution, two tails, $\alpha = 0.05$, $Ho : \mu_1 = \mu_2$

tails	α	z_{crit}
2	0.05	1.96

$$n = \frac{2\left(\left|z_{1-\frac{\alpha}{2}}\right| + z_{\beta}\right)^2}{d^2}$$

Power	d 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1	1.2	1.4
0.25	*	*	*	*	*	*	*	*	*	*	*
0.5	769	193	86	49	31	22	16	13	8	6	4
0.6	980	245	109	62	40	28	20	16	10	7	5
2/3	1144	286	128	72	46	32	24	18	12	8	6
0.7	1235	309	138	78	50	35	26	20	13	9	7
0.75	1389	348	155	87	56	39	29	22	14	10	8
0.8	1570	393	175	99	63	44	33	25	16	11	9
0.85	1796	449	200	113	72	50	37	29	18	13	10
0.9	2102	526	234	132	85	59	43	33	22	15	11
0.95	2599	650	289	163	104	73	54	41	26	19	14
0.99	3675	919	409	230	147	103	75	58	37	26	19

*   We are considering only values of beta greater than 0.5
A power value of 0.25 corresponds to a beta equal to 0.75

References

American Psychological Association (1994), *Publication Manual of the American Psychological Association*, 4th ed., Washington, DC: Author.

American Psychological Association (2001), *Publication Manual of the American Psychological Association*, 5th ed., Washington, DC: Author.

Berenson, M. L. and D. M. Levine (1996), *Basic Business Statistics*, 6th ed., Englewood Cliffs, New Jersey: Prentice Hall International Inc.

Berenson, M. L., D. M. Levine and T. C. Krehbiel (2004), *Basic Business Statistics*, 9th ed., Upper Saddle River, New Jersey: Pearson/Prentice Hall.

Bernstein, S. and R. Bernstein (1999), *Schaum's Outline Elements of Statistics II: Inferential Statistics*, New York: McGraw-Hill.

Borenstein, M. (2008), *Meta Analysis*, www.statistics.com online course.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd Ed., NewYork: Lawrence Erlbaum Associates.

Freund, J. E. (1962), *Mathematical Statistics*, Englewood Cliffs, New Jersey: Prentice Hall International Inc.

Kazmier, L. J. and N. F. Poll (1979), *Basic Statistics for Business and Economics*, 2nd ed., New York: McGraw-Hill.

Kline, R. B.(2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*, American Psychological Association.

Knüsel, L. (1998), "On the Accuracy of the Statistical Distributions in Microsoft Excel 97", Comput. Statist. Data Anal. 39, pp. 109-110 (With Corrigendum in CSDA, 40, 385).

Knüsel, L. (2002), "On the Reliability of Microsoft Excel XP for Statistical Purposes", Comput. Statist. Data Anal. 26, pp. 375-377.

Knüsel, L. (2005), *On the Accuracy of the Statistical Distributions in Microsoft Excel 2003*, Comput. Statist. Data Anal. 48, pp. 445-449.

Larson, R. and E. Faber (2000), *Elementary Statistics: Picturing the World*, Pearson Education.

Levin, R. I. and D. S. Rubin (1998), *Statistics for Management*, 7th ed., Upper saddle River, New Jersey: Prentice Hall International Inc.

Levine D. M., M. L. Berenson and David Stephan (1999), *Statistics for Managers Using Microsoft Excel*, 2nd ed., Upper saddle River, New Jersey: Prentice Hall International Inc.

Microsoft Corporation (2007), "Description of the effects of the improved statistical functions for the Analysis ToolPak in Excel", http://support.microsoft.com/kb/829208/en-us (accessed on on 2008-March-4).

Microsoft Corporation (2007a), "Floating-point arithmetic may give inaccurate results", http://support.microsoft.com/kb/78113 (accessed on 2008-March-4).

Microsoft Corporation (2007b), "How to correct rounding errors in floating-point arithmetic", http://support.microsoft.com/kb/214118/en-us (accessed on 2008-March-4).

Murteira, B., C. S. Ribeiro, J. A. Silva and C. Pimenta (2002), *Introdução à Estatística*, Lisboa: McGraw-Hill.

Pestana, D.D. and S. F. Velosa (2002), *Introdução à Probabilidade e Estatística*, Lisboa: Fundação Calouste Gulbenkian.

Pinto, J. C. C. and J. J. Dias Curto (1999), *Estatística para Economia e Gestão: Instrumentos de Apoio à Tomada de Decisão*, Lisboa: Sílabo.

Reis E., P. Melo, R. Andrade and T. Calapez (2001), *Estatística Aplicada* − Vol. 2, 4th ed., Lisboa: Sílabo.

Sanders D. H. (1990), *Statistics: A First Course*, 5th ed., New York: McGraw-Hill.

Sandy, R. (1990), *Statistics for Business and Economics*, New York: McGraw-Hill.

Sleeper, A. (2006), *Six Sigma Distribution Modelling*, New York: McGraw-Hill.

Smithson, M. (2000), *Statistics with Confidence*, London: SAGE.

Smithson, M. (2002), *Confidence Intervals*, Thousand Oaks, California: SAGE.

Spiegel, M. (1972), *Schaum's Outline of Statistics*, New York: McGraw-Hill.

Walkenbach, J. (2001). *Excel 2002 Formulas*, New York: M&T Books.

Walkenbach, J. (2008). http://www.j-walk.com/ss/excel/odd/index.htm.

Watson, C. J., P. Billingsley, D.J. Croft and D.Huntsberger (1993), *Statistics for Management and Economics*, 5th ed., Boston: Allyn and Bacon.

Webster, A.L. (1998). *Applied Statistics for Business and Economics*, 3rd ed. Boston: Irwin/McGrwaw-Hill.

Trade Marks and Intellectual Property

Microsoft® and Excel® are registered trademarks of Microsoft Corporation in the United States and in other countries.

SPSS® and Sample Size® are registered trademarks of SPSS Inc. in the United States and in other countries.

António Teixeira
UNIDE researcher
IBS – ISCTE Business School
Lisbon, Portugal
Phone: 351217903201
afst@iscte.pt

Álvaro Rosa
UNIDE researcher
IBS – ISCTE Business School
Lisbon, Portugal
Phone: 351217903201
alvaro.rosa@iscte.pt

Teresa Calapez
UNIDE researcher
IBS – ISCTE Business School
Lisbon, Portugal
Phone: 351217903232
teresa.calapez@iscte.pt