Methodology for testing statistical significance

Introduction

1. This annex sets out the approach taken to testing the robustness of certain results reported in the Audit Survey working paper, by reference to examples extracted from the main document.

2. We mostly rely on the significance testing provided by IFF Research applied to results for different company categories and to responses given by FD/CFOs versus ACCs (see IFF technical report). We carried out further significance testing where not provided by IFF Research.

3. The contents of this annex are as follows:

   (a) an overview of the data and how it is used;

   (b) the use and calculation of confidence intervals;

   (c) the application of the finite population correction; and

   (d) the calculation of significance tests.

Overview and use of data

4. A complete description of the survey data is reported in the Survey results working paper. In the working paper we define the population, the sample frame, and the achieved sample.[1] For the purposes of significance testing:

   (a) whilst the overall population of the survey is comprised of all the companies in all categories, the relevant population will vary depending on the results being analysed. Generically we denote the relevant population with N; and

[1] See paragraphs 5–7.

(b) whilst the achieved sample is comprised of all the companies in each category for which we have responses, as above, the relevant sample under consideration will vary and is generically denoted with *n*.

5. For instance, looking at FTSE 100 companies only, the population under consideration is composed of 100 companies ($N = 100$), and the achieved sample is 58 ($n = 58$).[2]

6. The minimum base size is a sample size threshold level under which a statistical analysis should not be performed because it is not sufficiently reliable.

7. In line with good practice, we report findings only when the base sample size is 30 observations or more. This figure is not imposed by statistical theory, but is based on judgement. Further information on data collection and minimum base size can be found in the IFF report.

Confidence intervals

8. A confidence interval is usually expressed as the estimated proportion, or mean, plus or minus its margin of error at a given confidence level. A confidence level indicates the probability that the true population parameter lies within the estimated interval, given a random sample extracted from the population. Whilst we did not report confidence intervals in the main paper we refer to these in paragraph 7 of the working paper, and below.

[2] This example is extracted from the working paper on CC survey results, Table 1, p3.

9. In the analysis of the survey results we used the 95 per cent confidence level in computing confidence intervals, ie that the true population parameter is expected to lie within the confidence intervals in 95 per cent of the cases.

10. Assuming that the sample is extracted from a population that has a Normal distribution, the formula for a confidence interval for sample sizes greater than 30 observations is the following:

$$\mathrm{x} - \mathrm{z}\frac{\mathrm{s}}{\sqrt{\mathrm{n}}} < \mu < x + z\frac{\mathrm{s}}{\sqrt{\mathrm{n}}}$$

Where μ is the (unknown) population parameter or characteristic of interest, x is the sample characteristic describing the relative population parameter, s is a measure of the variability of the sample (the so-called standard deviation), n is the sample size and z is a value corresponding to the level of confidence chosen for a Normal distribution. The term $\pm z\frac{s}{\sqrt{n}}$ is also known as the margin of error.

11. For a Normal distribution and a sample size of at least 30 observations, a 95 per cent level of confidence corresponds to a value of z equal to 1.96 ($z = \pm$ 1.96). For example, if 50 per cent of the respondents to a question in the survey state a characteristic of an audit firm as 'very important' at a confidence level of 95 per cent, with a margin of error equals to (plus or minus) 3.1 per cent, it means that if the survey were conducted 100 times, the percentage of the population that would list this characteristic as 'very important' will range between 46.9 and 53.1 per cent in 95 of the 100 surveys.[3]

Finite population correction

12. In order to compute confidence intervals, statistical theory assumes that the population is much (infinitely) greater than the sample. For finite populations and

[3] Using the above confidence interval notation: $0.5 - 0.031 < \mu < 0.5 + 0.031$.

samples greater than 5 per cent, a correction factor must be taken into account. In this case we have finite populations and achieved samples that are considerably more than 5 per cent of the population. We therefore applied a finite population correction (FPC).

13. The FPC, which takes a value between 0 and 1, is used as a multiplier that reduces the margin of error calculated for a random sample selected from a large population and so narrows the confidence intervals, ie it provides better estimates of the parameters of interest.

14. The FPC is computed using the information available on the population *N* and on the sample *n*:

$$FPC = \sqrt{\frac{N-n}{N-1}}$$

15. The formula of a confidence interval including a FPC is the following:

$$\mathrm{x} - \mathrm{z}\frac{\mathrm{s}}{\sqrt{\mathrm{n}}}\mathrm{FPC} < \mu < x + z\frac{\mathrm{s}}{\sqrt{\mathrm{n}}}\mathrm{FPC}$$

16. If we consider the example in paragraph 11—where 50 per cent of respondents rate a characteristic of an audit firm as ‘very important’ at a confidence level of 95 per cent, with a margin of error equal to (plus or minus) 3.1 per cent—using the information in Table 1 of the main document, the finite population correction for the company level sample of FTSE 100 companies is equal to 0.65 and the relative margin of error becomes approximately equal to 2.0, reducing the spread of the confidence interval.[4] This implies that, given the sample characteristics, in 95 per

[4] The information used to compute the FPC in this example is extracted by Table 1: *N* = 100; *n* = 58. In this case this leads to a margin of error of 2.015.

cent of the cases a number of FTSE 100 companies between 48.0 and 52.0 is likely to assess this specific audit firm characteristic as 'very important'.

Statistical significance testing

17. In the main document we comment on how responses differ between mutually exclusive categories, such as, for instance, how responses for FTSE 350 companies compare with those for other company categories.[5] For all such comparisons we reported results only where these are statistically significant (see paragraph 9). These are denoted with a *.

18. A statistical significance test involves the choice of a null hypothesis—a statement regarding the survey findings that aims to explain a result, usually denoted as H_0– and of an alternative hypothesis H_A, that negates such explanation. The two hypotheses are therefore stated in such a way as to be mutually exclusive. The test allows us to either accept the null hypothesis H_0, or to reject it in favour of H_A.

19. By testing hypotheses at 95 per cent significance level, results can be considered to be statistically significant given that there is only a 5 per cent chance (confidence level) that the null hypothesis is rejected by coincidence instead of because of a pattern in the observations.

20. We used a (two-tailed), two-proportions Z-test to analyse company level results to assess whether observed differences in responses for different company categories are significantly different.

---

[5] For example, in paragraph 13*(b)* of the main document we analyzed the audit tenure length together with audit fee.

The two-proportions Z-test

21. To apply the two-proportions Z-test the following conditions must be met: samples for each population are selected using simple random sampling; the samples are independent, i.e. there is no relation between them; and each sample size n is greater than or equal to 30 observations. In this case the achieved sample is not a random sample of the population. Nevertheless for the purposes of carrying out significance testing we assume this to be the case. In particular, we assume that there is not a systematic bias in the sample.

22. For the two-proportions Z-test the null hypothesis is that there is no difference between two groups with respect to a specific characteristic and the alternative hypothesis is that the observed percentages for the two groups relating to the specific characteristic are statistically different (H_A):

$$\begin{cases} H_0: p_A = p_B \\ H_A: p_A \neq p_B \end{cases}$$

23. This particular formulation of the alternative hypothesis H_A denotes that the test is a two-tailed test, meaning that an extreme value on either side of the sampling distribution would imply the rejection of the null hypothesis.[6] Therefore the test rejects the null hypothesis if the proportion for one group of respondents giving a particular response is significantly higher or lower for than for another group.

24. The inputs to the calculation of two-proportions Z-test are:

(a) the proportion or percentage of each group of respondents giving a particular response to a survey question (eg those who said that the experience of the audit partner was an important consideration in the appointment of the auditor). We denote these percentages as p_A and p_B respectively;

[6] We do not implement one-tailed tests because we are not interested in knowing if a group of respondents has a greater or smaller percentage of agreement with a survey statement than another group.

(b) the respective achieved sample sizes for these two groups, n_A and n_B, ie the total number of respondents who were asked a particular question (eg the total number of individuals that have been asked what characteristic is important in the appointment of the auditor).

25. Using the observed sample characteristics, the value of the Z-test is then simply the ratio between the percentages' difference and the standard error:

$$z = \frac{p_A - p_B}{SE}$$

26. To compute the standard error *SE*:

(a) First compute a pooled sample proportion *p* as a measure of the weighted sample average of the proportions of the two groups:

$$p = \frac{p_A n_A + p_B n_B}{n_A + n_B}$$

(b) Then, compute the standard error *SE*, applying an FPC to each group is given by:

$$SE = \sqrt{p(1-p)[\frac{FPC_A}{n_A} + \frac{FPC_B}{n_B}]}$$

27. A value of z is obtained by applying the formula given in paragraph 24 above. This must then be compared with the theoretical critical value corresponding to a 95 per cent level of confidence, that is z = ± 1.96. The null hypothesis is rejected whenever the estimated z value exceeds the theoretical z threshold in absolute value.

28. For example, let us consider question B3 and focus on the differences between FTSE 350 versus all other companies (following the above mentioned notation, (respectively group A and B). The strength of international network is an important characteristic in deciding to appoint or reappoint a statutory auditor for 64 per cent

(p_A) of FTSE 350 companies (with question sample base n_A equals 195 companies) and for 45 per cent (p_B) of all other companies (with corresponding sample base n_B equal to 279 companies).

29. In this example, the pooled sample proportion *p* is equal to 0.52. The FPC for each group can be easily computed by means of sample and population sizes in Table 1. Combining these elements, the estimated standard error *SE* is equal to 0.034. Then the z value of the observed data is equal to 5.6,[7] much greater than the 95 per cent corresponding confidence level, z = 1.96.

30. This means that we reject the null hypothesis stating that the percentages coming from the two samples are equal, meaning that there is a statistically significant difference between the two groups concerning auditor characteristic.

31. Conversely, if the null hypothesis of equal percentages would have been accepted because the z value was found to be less than 1.96, then we would have claimed that there is no difference between the two groups with respect to the strength of international network of the statutory auditor in deciding a (re-)appointment.

[7] The calculation used to obtain the Z value, based on that given in paragraph 24, is Z = (0.64–0.45)/0.034 = 5.58.