

POPULATION



ANALYSIS

AN EASIER WAY TO IDENTIFY MEDIAN EMPLOYEE PAY

By Nathan O'Connor and Josh Schaeffer, Equity Methods LLC

It's time to look at a new way to find answers.

“Dewey Defeats Truman”
Chicago Daily Tribune,
Nov. 3, 1948.

Arguably the most famous newspaper gaffe in history, this error is frequently attributed to sampling bias caused by an improper sampling technique.

It's a technique that is drawing more attention these days.

In the recently released rule on pay ratio disclosures, the U.S. Securities Exchange Commission (SEC) provided flexibility in determining the compensation rate of the median employee. Companies can select from various

measures of compensation, such as total compensation or other metrics based on payroll data, as long as they are consistently applied across all employees. Companies also can select the median employee from all employees or a statistical sample. This sampling provision in the final rule has triggered a flurry of debate among experts about the variety of statistical sampling techniques that can be deployed to identify the median employee.

Equity Methods LLC works with clients who are considering using sampling instead of the full



A key issue with sampling is that analyzing multiple samples, even with the same methodology, will result in different estimates of the median employee each time a sample is drawn.

population, but so far, our research shows that the costs of sampling will frequently outweigh the benefits. The company concludes that based on a seemingly simple question that has remained largely unaddressed to date: If the company chooses to identify its median employee by using a sampling technique, and to form its sample, it will need to identify a population from which to sample. Why not choose to identify the median employee from the entire population itself?

Equity Methods shows the advantages and disadvantages of sampling in this article and compares it with the costs and benefits of analyzing the full population. The advantages of sampling may appeal to some companies, but in many cases, these advantages may not pan out in the actual implementation.

Why Sample in the First Place?

Before we walk through some limitations of statistical sampling, it may help to clarify what statistical sampling is and why the SEC decided to allow companies to use it. Statistical sampling uses observations from a relatively small, randomly selected sample of data points to draw conclusions about a broader population. In this case, it amounts to using a randomly selected set of employees to draw conclusions about all employees at a corporation. The advantages of sampling are that it

limits data-gathering costs and it allows experimenters to work with more manageable datasets. When data are costly or otherwise hard to obtain, a well-designed sample can be crucial to developing any conclusions at all.

Potential Issues with Statistical Sampling

The primary benefit of sampling — cost savings — is offset by several potential issues.

Sampling Results Can Vary

A key issue with sampling is that analyzing multiple samples, even with the same methodology, will result in different estimates of the median employee each time a sample is drawn. For example, a company with 1,000 employees may attempt to draw five samples from 100 employees to test which employee should be selected and the stability of the selection process. It would not be surprising if the selected median employees from each sample were employees ranked 560, 478, 402, 521 and 489 in terms of pay scale while the firm employing the population would always select the employee ranked 500. While for the “average” sample, the selected employee will be very close to the median, the result from any individual sample may be further away. The size of this difference is related to the variability of the distribution and

the size of the sample selected. In an earlier blog post, we attempt to quantify how big the impact of these differences can be.

Of course, a company that develops a pay ratio estimate that is either above or below the median from a full population analysis exposes itself to claims that the number was strategically selected. Although the company, ultimately, may be able to show that the sample was appropriately built, the level of complexity in these analyses may result in excessive burdens or costs for companies asked to defend their methods.

Moreover, the variance of sampling will cause multiple potential problems for companies:

- Companies have no way to know how the median from their selected sample compares with that of the whole company;
- Selecting different median employees over time may result in a shift in the calculated pay ratio only because of the random differences from different samples;
- If the demographics of the employee population change between selections of median employees, the company will need to update its sampling methodology.

All of these problems can be mitigated by using larger samples, but as the sample size increases, any cost benefits of sampling over using the full population are diminished.

Challenges with Selecting Specific Groups

One situation where sampling can be of great benefit is when sampling from a subset of a larger group. For example, a fast food restaurant testing a new burger can sample potential customers only in Ohio because Ohio is known to be a state with similar population demographics and tastes to the United States. It would be silly to select a sample of people in Hawaii for the test if the demographics don't line up to the larger customer base. In statistics parlance, the subset is called the sampling "frame," and samplers need to make sure that the frame is representative of the overall population.

Unfortunately, companies likely will be unable to make conclusions based on a subset. For example, companies who wish only to look at their U.S. employees would have to conclude that foreign employees

have the same general pay. Given the difference in the roles of foreign and U.S. employees as well as economic differences among countries, this assumption might be problematic. Considering a case where a company did use only U.S. employees, if foreign employees had lower pay, eliminating the foreign data would bias the median pay up and the ratio down. This can be solved through an analysis of the populations; however, that may defeat the cost-saving purpose of sampling.

Why Population Is Required

Even when using statistical sampling, it is still necessary to 1) identify the employees in the sampling frame and 2) collect information for those employees. Because of the representation issues, savings from not needing to query data from multiple databases may not materialize, and companies will still need to get data they may

have wished to avoid from foreign and consolidated subsidiaries.

Difficulty with Determining an Appropriate Sample Size, and Sample Sizes May Be Large

The SEC has chimed in on the crucial question of how big of a sample companies will need. In the final rule release, the SEC recounted its experiments on what sample size would be needed for companies in various industries and found required sample sizes larger than many expected. Using a set of simplifying assumptions, such as a company with only one segment and geography, the lowest estimated sample size needed was 81, while the highest was more than 1,500 based on statistics at an industry level. A company can do a pilot analysis to determine what an appropriate sample size would be, but a pilot analysis can be costly.



● ● ● using the full population delivers benefits in the form of certainty that the number is right and supportable, as well as opportunities for automation and process optimization.

Are Data Acquisition and Analysis Expensive?

Given the concerns we see with sampling, we now consider the primary perceived benefits: lower cost of data acquisition and the use of smaller datasets. If a detailed calculation is needed for each employee selected, the marginal cost of adding an additional employee to the pool could be substantial. However, because companies can use payroll data and because any modifications such as cost of living would be consistently applied across employees and would be determined for all employees in the sample, the effort to do the pay ratio analysis for a sample or for the entire population will be very similar, if not identical.

Concerns abound that performing population analysis is not feasible because of the large datasets involved. First, the SEC estimates that more than half of registrants will require a sample size of more than 500 employees, which will likely offer limited or no savings over using the employee population. In fact, costs may be higher when a sample is used because it will be tougher to automate processes. Second, procedures exist for extracting, merging and comparing data from multiple payroll systems — it's not always easy, but that nut has been cracked time and time again. Third, modern computer technology and the advent of big data help solve this problem. For example, we routinely analyze many gigabytes of new data received each month from clients, and for far more complex purposes than pay ratio calculations.

Benefits of Population Analysis

As companies will likely gain little advantage on the two main sampling fronts, we note that there may be little or no advantage to using a sample, despite the inflated cost of sample design and implementation. On the other hand, we note that companies using a population may see several benefits from this analysis.

First, companies doing this analysis will have access to a full complement of payroll data from all of their operations. Such data may be useful for internal budgeting, allocations and planning, among other purposes.

Second, companies calculating their pay ratio with a comprehensive analysis will be able to more accurately compare the impact of different compensation measures, such as base pay and bonus. Also, cost of living adjustments or other allowable options will be measured without worrying that any difference is because of characteristics of the selected sample.

Third, companies moving between median employees after three years, when the rule requires a new median be calculated, will not have to worry about shifts because of merely selecting a different sample and thus will have more comparable pay ratio numbers between years.

Choosing a Method That Works

With the release of the final pay ratio rule, companies will soon be making difficult decisions about how to identify their median employee. Companies have an important choice before them: select a sample from their employee base and identify the pay of each of

those employees, or determine their actual median employee from a population of all employees.

The statistical bar for doing sampling right and meeting the SEC's expectation that a "reasonable statistical sample" be selected is relatively high — high enough that most of the legwork involved in pulling from the full population needs to be expended, and companies should be prepared to pay for experts to help with designing their sample. But using the full population delivers benefits in the form of certainty that the number is right and supportable, as well as opportunities for automation and process optimization. As a result, we expect to see companies spend their time focused on the potentially more fruitful exercise of aligning and combining payroll systems than developing a potentially problematic sample. **WS**

Nathan O'Connor is managing director at Equity Methods LLC in Scottsdale, Ariz. Contact him at nathan.oconnor@equitymethods.com.

Josh Schaeffer is director of Complex Securities at Equity Methods LLC in Scottsdale, Ariz. Contact him at josh.schaeffer@equitymethods.com.

resources plus

For more information, books and education related to this topic, log on to www.worldatwork.org and use any or all of these keywords:

- Compensation
- Data Analysis
- Pay Ratio