Statistical Pay Equity Analyses:
Data and Methodological Overview

by
Paul F. White, Ph.D.
Resolution Economics, LLC
Washington, DC

for
American Bar Association
10th Annual Labor and Employment Law Conference
Chicago, Illinois

November 9-12, 2016

1. Introduction

Many, if not most, employers have written policies that prohibit discrimination against any employee on the basis of a protected group status with respect to employment decisions such as compensation, hiring, promotions, performance evaluations, and terminations. However, these policies do not specify how potential discrimination should be identified and corrected. There are a variety of different methods, each with their pros and cons, by which potential discrimination can be identified and quantified within in a workforce. This leads to uncertainty for employers and employees with respect to the best way to detect, monitor, and correct potential discrimination.

For decades, employers, employees, and the Courts have relied upon the analysis of employee data to determine whether disparities adverse to a protected group of employees exist within a workforce. For example, is it the case that females receive less compensation than their similarly-situated male colleagues? The answer to questions such as these can have an important impact on employers' policies and practices, legal exposure, and profitability.

The purpose of this paper is to provide an overview of the data and statistical issues that may need to be considered when conducting statistical analyses of potential compensation discrimination within a workforce. This paper discusses gender disparities in compensation, but the same concepts may be applied to potential discrimination against other protected groups as well as to other employment decisions, such as hiring, promotions, etc.

2. Statistical Analyses – Initial Points

It should first be noted that statistical analyses cannot prove the absence or presence of discrimination; however, statistical analysis can be useful to support or rebut allegations of discrimination, or to show that a protected group has been affected by employment practices in a manner different from a non-protected group of employees. As such, data and statistical analysis may provide evidence to assist the finder of fact determine whether discrimination exists, where it exists, and the magnitude of any disparities, keeping in mind that statistical models, by their very nature, only incorporate information contained in the data upon which they are based. Due to the fact that some determinants of compensation cannot be quantified, and other factors are not maintained in a database, statistical analyses often do not account for everything that affects the decisions. As such, statistical findings need to be considered as one among several sources of information before determining whether discrimination exists and what remedial measures are to be undertaken.

A general concept in the statistical analyses of alleged discrimination is the concept of comparing what is actually happening to what would be expected to happen in a neutral decision-making process. For example, when examining compensation among a group of similarly-situated employees, one would expect the compensation of females to be approximately the same as their male counterparts. If the actual compensation of females is different than males (either higher or lower), then the statistical tests will determine whether the differences could have occurred by random chance, or whether the difference between actual and expected compensation is statistically significant and not likely to have occurred by random chance.

3. Comparisons of Averages

A common calculation at the early stages of an analysis is a simple comparison of the average compensation of the protected employees to the average compensation of the non-protected employees. This is sometimes done at the company-wide level or at some other organizational level (e.g., pay grade or job group). In fact, the popular press often reports the gender pay gap by reporting the average or median earnings of females among all workers in the labor force compared to all working males.

For example, in the below table, across all pay grades there is a $3,426 difference in the average salary levels between males and females (representing a pay gap of almost 2%). However, the direction and size of this relationship is different for each pay grade.

Grade	**Average Female Salary**	**Average Male Salary**	**Difference**
5	$118,499	$118,043	$456
6	$116,224	$117,027	-$803
7	$129,640	$132,623	-$2,983
8	$122,222	$120,060	$2,162
9	$116,947	$125,078	-$8,131
10	$117,347	$123,851	-$6,504
Overall	**$122,345**	**$125,771**	**-$3,426**

Relying solely on averages is usually not informative for a number of reasons. First, unless the compensation policies and practices are rigidly structured, there are many factors that determine compensation levels. These factors, which will be discussed later, are usually not considered when calculating broad averages, thus making comparisons potentially misleading.

Another reason why a comparison of averages may be misleading is that it may mix employees who are not similarly situated, due to organizational or individual differences which should be taken into account. Especially for medium-sized and large employers, there are often multiple different compensation structures within the same workforce. Combining these structures into the same average could result in either false positive or false negative findings.

Furthermore, by definition, averages may be influenced by outlier individuals whose compensation levels are either unusually high or unusually low compared to the others in the analyses. There are many reasons why this can happen (e.g., lateral hire, specialized skills, incorrect data, etc.), These outliers may be sufficiently different that they have an outsized effect on the average for their group.

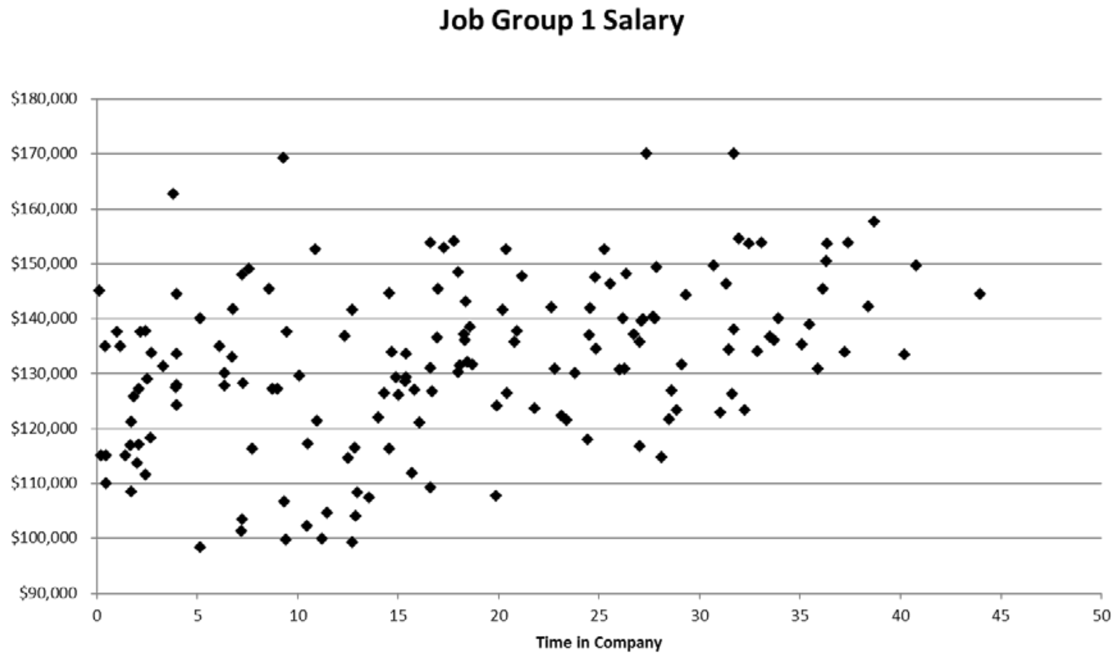
4. Regression Analysis

As mentioned above, there can be many factors that affect the compensation levels of employees. Some are related to their place in the organization (e.g., pay grade, department, division, location, reporting manager, etc.) and some factors are related to their individual characteristics (e.g., time in job, time in grade, company seniority, performance, education, training, etc.). These factors vary from employer to employer, and usually within an employer.

A multiple regression analysis is a statistical procedure that accounts for the differences between protected and non-protected employees with respect to these “control factors” (i.e., organizational and individual characteristics). The regression then measures the compensation differences between protected and non-protected employees after accounting for their differences among the control factors.

To illustrate, suppose the average difference in compensation between males and females was $10,000 in favor of males. Is that a sign of discrimination? Suppose males had on average 10 more years of seniority than females. Thus, accounting (or “controlling”) for company seniority may be sufficient to explain the difference in average compensation between males and females. In this example, a regression analysis is a statistical procedure that determines how much of the male/female pay gap is explained by seniority.

Another way to think about a regression analysis is using a scatterplot. In the below chart, we see the relationship between company seniority and salary. Each data point represents an employee.



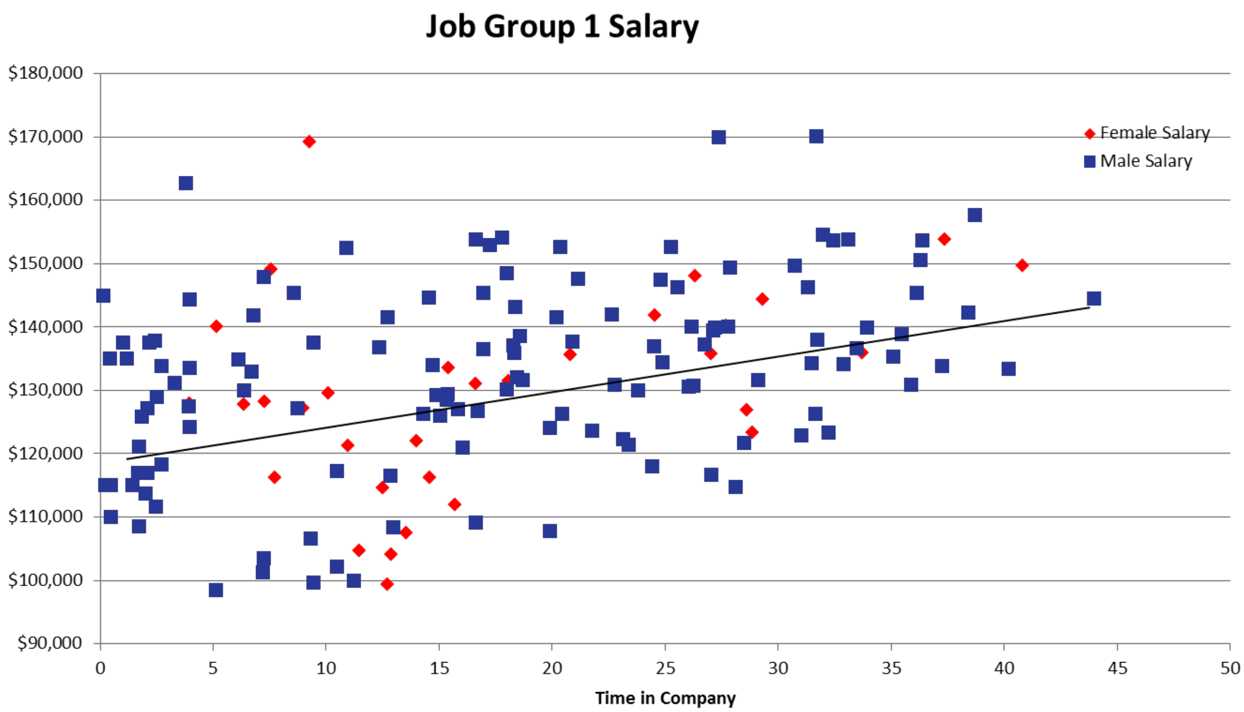
We can see from the above chart that there is a positive relationship between company seniority and salary, meaning that more years of seniority are generally associated with a higher salary. This is not a perfect relationship because there are other factors that affect salary and are not included in this chart.

The regression can be thought of as measuring a trend line that best fits this data. In the below chart, the trend line confirms the positive relationship between seniority and salary.

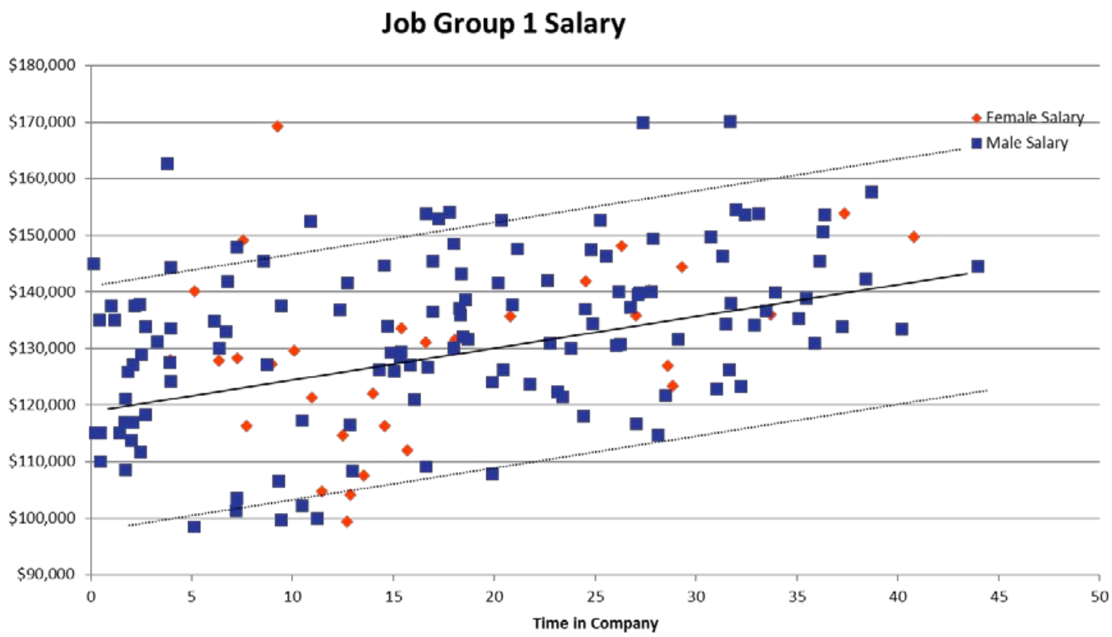
Going back to the earlier discussion about "actual" and "expected" compensation, the trend line can be thought of as the expected level of compensation for each individual (based on the model used), and the dots indicate actual compensation. Thus, anyone above the trend line is paid more than expected, and anyone below the trend line is paid less than expected. The vertical distance between each dot and the trend line represents the amount by which people are paid above or below their expected values.

In this example, if compensation decisions were gender-neutral, one would expect to find both males and females above the trend line, as well as below the trend line. Employers whose female employees are clustered below the trend line and males above the line are often faced with a statistically significant pay disparity. (Keep in mind, any analysis needs to be restricted to employees who are similarly-situated).

The below chart shows the relationship between company seniority and salary, differentiated between males and females. As can be seen from the chart, males and females are found both above and below the trend line, and there are instances where both males and females have observations that are inconsistent with the rest of the data (i.e., combinations of salary and seniority that are sufficiently distant from the trend line).



A multiple regression analysis provides information on which employees are paid statistically significantly above or below their expected values. A common measure by which statistical significance is determined is the "standard deviation". A difference between actual and expected compensation that is more than 1.96 standard deviations would have less than a 5% probability of occurring by random chance, and is therefore thought of by most courts as statistically significant. The below chart shows the statistical significance ranges around the trend line in our compensation example. Employees above the top dashed line are paid statistically significantly higher than expected given their level of seniority. Similarly, those below the bottom dashed line are paid statistically significantly less than expected given their company seniority.



As before, one would expect an even distribution of males and females paid significantly more and significantly less than expected. Employers who tend to have females clustered below the bottom dashed line and males above the top dashed line are more at risk of having a statistically significant gender pay disparity somewhere in the company.

These outlier employees may warrant further investigation to determine whether their data is correct, whether they are similarly situated to the others in the analysis, or whether there are perhaps additional factors influencing their compensation, not previously accounted for, that could help explain why they appear to be outliers.

5. Methodological Considerations

From a methodological standpoint, there are a number of decisions that could be made that may have an important effect on the model structure and thus the findings. These decisions include:

a. Whether to perform a single regression for the entire company, versus conducting a separate regression for sub-populations of the workforce,

b. Which control factors to include in the regression model, and

c. Whether to customize the structure of the model to address the requirements of a specific law or regulation.

6. Not all Groups of Employees are Suitable for Regression Analysis

There are some instances when there are too few people in the population or sub-population to conduct a regression analysis. In those instances, other statistical tests may be used to provide insight into whether there are disparities that need to be addressed. Two commonly-used approaches are the Rank Sum test and a "cohort" analysis.

A <u>Rank Sum test</u> measures whether there is a statistically significant difference in the distribution of compensation between the protected and non-protected employees. Here are the basic steps the Rank Sum test takes:

a. Sort the employees in the analysis in ascending order of salary (i.e., those with the lowest salary are at the top of the list and those with the highest salary are at the bottom of the list).

b. Assign a rank of 1 to the person with the lowest salary, 2 the next-lowest salary, etc.

c. If there are people who have the same salary, their rank is the midpoint of the rank levels they share. For example if the 4^{th} and 5^{th} ranked people have the same salary, both would receive a rank of 4.5 (the midpoint between their respective levels).

Once the list has been sorted and the ranks have been established, the Rank Sum test determines whether the protected employees are evenly dispersed throughout the list, or are clustered among the low or high-salaried employees. The Rank Sum test compares the average rank of the protected and non-protected employees and determines whether the difference between the two is statistically significant. Below is an illustration:

Name	Gender	Salary	Rank Score
Kenny	M	$20,000	1
Aaron	M	$85,000	2
Nikki	F	$87,000	3
Emma	F	$90,000	5.5
Miah	F	$90,000	5.5
Thea	F	$90,000	5.5
Theo	M	$90,000	5.5
Andy	M	$91,000	8
Amelia	F	$92,000	9
Noah	M	$92,500	10
Tony	M	$93,000	11
Walter	M	$93,500	12
James	M	$94,000	13
Lex	M	$95,000	14
Capi	M	$98,000	16
Dimi	M	$98,000	16
Theron	M	$98,000	16
Wren	F	$99,000	18
Janet	F	$100,000	19.5
Mark	M	$100,000	19.5
Dan	M	$111,000	21
Bob	M	$115,000	22
Jen	F	$120,000	23

	Female	Male
Count	8	15
Average Salary	$96,000	$91,600
Average Rank Score	11.13	12.47

In the above example, the average rank for females is 11.13 compared to the average rank of males of 12.47 (recall that because the list is sorted from lowest to highest, a low rank is associated with a low salary). However, this difference is not statistically significant, which indicates that the differences could have reasonably occurred by random chance.

In this example, the Rank Sum test results and the average salaries yield contrary conclusions. The Rank Sum test shows that females have on average a lower salary rank, yet females have on average a higher salary level. This is driven primarily by the low-side salary outlier, who is a male with a $20,000 salary level. One would need to further examine this person to see if either his salary data is incorrect or whether he is really similarly-situated to the others in the analysis (and if not, perhaps should be excluded from this test).

It should be noted that while the Rank Sum test is valid in small sample sizes, it does not have the same ability as a regression to add further control factors (such as seniority or performance). The criteria used to group employees represent the only controls found in the analysis. For example, suppose the employee data is divided into groups of people who are in the same job group and pay grade. All of the job group/grade combinations are large enough to run a regression except for Engineers in Grade 1, then a Rank Sum test on Grade 1 Engineers is controlling for job group and grade only. It does not have further controls for seniority, performance, etc.

A "cohort" analysis can also be useful when examining the compensation levels for a small number of people. While this analysis can take on many different forms, the cohort analysis does not involve tests of statistical significance. Rather it addresses the question of whether it appears that the compensation differences between protected employees and non-protected employees can be explained by differences in their other characteristics, such as seniority, job title, etc. In this regard the cohort analysis is the same as the regression, but the cohort analysis does not need to rely on statistical tests. The below table is an illustration of two individuals who may be thought to be similarly-situated:

	Person A	**Person B**
Gender	Male	Female
Job Description	Engineer II	Engineer II
Salary Grade	8	8
Job Function	Widgets	Widgets
Area Differential	5%	5%
FT/PT	FT	FT
Exempt Status	Ex	Ex
LOA	No	No
Recently Demoted	No	No
Recently Promoted	No	Yes
Performance Eval.	None	Outstanding
Years in Current Grade	0.10	0.75
Years in Other Grades	0	21.57
Years Pre-Company	22.00	1.0
# Direct Reports	0	2
Rehired	No	No
Salary	$95,000	$75,000

In the above example, the male (Person A) is very similar to the female (Person B) except for the fact that the male was hired directly into the Engineer II position, whereas the female had spent many years at the company before she was eventually promoted into the position. They both have similar levels of relevant experience prior to moving into the job, but the male's experience was with another company, whereas almost all of the female's career was spent with this company.

The question to be answered in this type of paired comparison is whether the male's experience prior to joining the company is sufficient to justify the $20,000 difference in salary (despite being in the same job and having a similar amount of relevant experience prior to joining the company). Or is it the case that there are other characteristics of these two people that are not included in the table but explain the difference in salary levels? It should also be noted that cohort analysis comparisons do not need to be solely between two individuals. Rather it could be a comparison between an individual and multiple similarly-situated cohort members. The key consideration is the methodology behind defining the group of people believed to be similarly-situated.

7. Data Issues

The analyses discussed above are built on the foundation that the underlying employee data is correct. However, it is not uncommon for data to be inconsistent, incorrect, or incomplete. Before conducting the statistical tests, the analyst should review the underlying data to determine the extent to which these issues exist. Below are some examples of common data issues that may affect the outcome of statistical analyses if not addressed.

a. Missing demographic information: Some employees choose not to self-identify their demographic information.

b. Inconsistent dates: Employee data files usually contain important dates such as date of hire, date entering current job, date entering current grade, leave of absences, merit awards, promotions, date of birth, etc. It is not unusual to observe inconsistencies between these various dates (e.g. date of hire is later than date entering current job).

c. Inconsistencies between job title and job family (e.g., an Engineer in the Human Resources job family).

d. Job titles that appear to be the same, but are recorded differently (e.g., "Electrical Engineer II" and "EE II"). First the analyst should work with the employer to confirm whether the various spellings actually represent the same job. If so, then an effort needs to be made to make the labeling consistent throughout the data before conducting the analyses. Otherwise, the statistical test will treat these people as if they are in different jobs.

e. Standard hours: If a workforce has people who are paid based upon something other than the standard 40-hour week, the data needs to include information on the number of standard hours that form the basis for their compensation levels so the analyst can prorate accordingly.

f. Legacy data systems: Sometimes employers who switch HRIS data systems do not maintain the entire work history data of their employees at the time of conversion to the new system. Other times the employers keep only the data for those who are employed at the time of the conversion. Even if the full information is maintained from the old system, there may be issues with respect to the consistency of data fields between the two systems that need to be understood.

8. Summary and Conclusions

The purpose of this paper is to provide an overview of the data and statistical issues that may need to be considered when conducting statistical analyses of potential compensation discrimination within a workforce. Regression analyses, as well as other statistical tools, provide insight into the question of whether there are statistically significant pay disparities between protected and non-protected groups of employees. However, the methodology behind the statistical tests is crucial in the outcome of the analysis. The decision of whether, and how, to subdivide the data and the control factors to use in the model directly affect the conclusions and subsequent remedial measures to address any disparities.