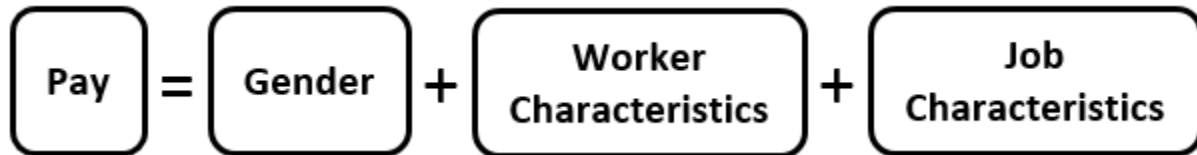


## Measuring Pay Equity--Large Sample Size Group Analysis

When sample size is sufficiently large ([see XXX reference intro](#)), multiple linear regression (MLR) is an analytical method commonly used by professional analysts. MLR is used to measure pay equity *controlling for* bona fide job-related factors, which can be Worker Characteristic (e.g., seniority, performance, education) or Job Characteristic (e.g., shift, hazard). The following figure summarizes the MLR approach to measuring gender pay equity:



Where,

- **Gender** = Estimated “effect” of gender on pay. It is a measure of pay equity.
- **Worker Characteristics** = Estimate “effects” of bona fide individual characteristics (e.g., seniority, performance, education) that may explain pay differences among individuals performing substantially similar work.
- **Job Characteristics** = Estimate “effects” of bona fide job characteristics (e.g., shift differential, hazard) that may explain pay differences among individuals performing substantially similar work.

### Using Multiple Regression Analysis to Test for Gender Disparities in Pay

Although conducting an MLR analysis requires specific expertise, it is possible for an HR professional to gain a conceptual understanding of the application of MLR to measure pay equity. Please note that the following is a simplified description of the steps involved in an MLR compensation analysis.

Step 1: Identify Employees in Substantially Similar Work Categories.

The California Fair Pay Act (FPA) was enacted to ensure that individuals performing substantially similar work are compensated equally. It is important to compare pay among individuals who are performing **substantially similar work** ([LINK TO STEP BY STEP](#)).

Step 2: Specify the Compensation Model Components

One of the most important steps in a pay equity investigation is to invest the time to understand the factors and forces that impact pay decisions. This is particularly important, because the FPA is focused on pay differences *after* bona-fide job related factors are accounted for. **Bona fide factors include** ~~Factors such as~~ tenure (time in company, and time in job), **and** performance, but this list is neither universal nor exhaustive. In fact, pay decisions may differ among jobs even if they are in the *same* company. Therefore, it is critical to understand the factors that impact pay decisions and to properly model them in the MLR analysis.

Step 3: Assemble the Data for Analysis

Once the factors in your compensation model have been specified, the next step is to assemble the necessary data for MLR analysis. For example, if Time in Job and Performance were identified in Step 2 as critical to the compensation model, then it is important to include reliable

## Measuring Pay Equity--Large Sample Size Group Analysis

measures of these compensation components for each employee in the analysis data file. Suggestions for data collection are discussed at [\[LINK TO DATA TOOL\]](#).

### Step 4: Prepare the Data for Analysis

There are two major types of data that may be included in the analysis: 1) Quantitative or Numeric and 2) Qualitative or Categorical:

- 1) Quantitative data measure the quantity of a characteristic (e.g., years of job tenure) and can take any numeric value. Some quantitative data can naturally take on numeric value, e.g., tenure is measured in years. In other instances, quantitative data can capture relational differences (e.g., more than, less than) but the data may not naturally be in numeric format. For example, performance data can contain ratings that range from low to high and these ratings can be coded out into numeric quantitative data format (e.g., high=3, med=2, low=1).
- 2) Qualitative/categorical data measure the presence of specific qualities or characteristics (e.g., gender, race, work location). These data require special treatment before they can be entered into a regression analysis. One of the more common methods of “transforming” categorical data into analyzable form is to *dummy code* them into a set of 0 and 1 indicator variables. For example, to dummy code gender, recode Female records into 0's and Male records into 1's. When a qualitative data contains more than two categories, a series of indicator variables are used to identify each category of the variable (rather than assigning sequential numeric values to each category, which is not appropriate for MLR analysis and will provide inaccurate analytic results).

### Step 5: Model Evaluation

Once the data are collected and prepared for analysis, the next step is to examine the relationship between each explanatory factor and pay to ensure that the final regression model is valid. There are many approaches to model evaluation and there is no one correct methodology. Here are some common analyses for model evaluation:

- 1) Examine whether each explanatory factor is significantly related to pay. Correlation analyses are helpful with quantitative/numeric data. More advanced methods of computing and evaluating change in  $R^2$  can evaluate both quantitative/numeric and qualitative/categorical data.
- 2) Examine the nature of relationship between each explanatory factor and pay (e.g., positively related, or negatively related). For example, higher performance should be related to higher pay.

For an MLR analysis to be reliable, it is important evaluate each explanatory factor to ensure that they significantly and appropriately related to pay.

### Step 6: MLR Analysis

The next step is to conduct an MLR analysis that includes all of the relevant bona fide factors and demographic characteristics of the employees. There are many statistical programs that can accomplish this but the general specifications are:

- 1) Compensation/Salary data is entered as the *Dependent Variable*
- 2) Gender (coded as an indicator variable) is entered as an *Independent Variable*

## Measuring Pay Equity--Large Sample Size Group Analysis

- 3) All bona fide characteristics (both quantitative and qualitative/categorical are entered as *Independent Variables*

### Step 7: MLR Results Interpretation

Interpreting MLR results involves the following steps:

- 1) Determine if each factor included in the model is meaningfully related to pay using the statistical test of the estimated regression coefficient (e.g., Z-test or t-test results for each regression coefficient generated by the statistical software). Each regression coefficient estimates the change in level of pay that is associated with a 1-unit change in the value of the explanatory variable. For example, if employee tenure is coded as years of employment, the estimated regression coefficient for employee tenure indicates the pay difference (usually an increase or “premium”) associated with an additional year of employment at the company. Explanatory factors that are not statistically significant are interpreted as having no impact on pay in the context of controls for the other explanatory factors included in the model; they should be discarded.
- 2) Determine if the Gender/Race factor is statistically significant. The estimated regression coefficient for the variable measuring employee Gender/Race accomplishes the primary goal of the pay equity analysis – to test if there is a significant gap in pay by gender/race, controlling for other bona fide factors associated with pay.
- 3) If the estimated coefficient for gender/race is statistically significant, the next step is to determine the direction and magnitude of the gender/race difference in pay. The sign (whether the value is positive or negative) indicates which group is paid more and the numerical value of the regression coefficient indicates the size of the pay gap. For example, if the gender variable is coded 0 for females and 1 for males:
  - a positive coefficient indicates that men are paid more than women, while a negative coefficient indicates that men are paid less than women (controlling for all bona fide factors included in the MLR analysis that might generate pay differentials between male and female employees).
  - The numerical value of the regression coefficient is an estimate of the size of the gender pay gap (after account for all bona fide factors included in the MLR analysis that might generate pay differentials between male and female employees).

This document provides an outline of a commonly-used approach to measuring pay equity when sample size, i.e., number of employees, is sufficiently large to support a Multiple Linear Regression analysis. Appropriate and accurate statistical analysis of pay equity require analytical expertise and experience which this report does not provide. Employers and HR professionals interested in conducting a pay equity analysis are strongly encouraged to seek expert consultation and/or services.