

# **Compare<sup>®</sup> 2.0**

## **Compensation Analysis Software**

**Technical User Manual**

**Version 11**

**Biddle Consulting Group, Inc.**

Copyright © 2011 Biddle Consulting Group, Inc.

Biddle Consulting Group, Inc.

193 Blue Ravine Road, Suites 250/270

Folsom, Ca. 95630

1.800.999.0438

All rights reserved. No part of this manual may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the publisher.



# Table of Contents

<b>I.</b>	<b>Conducting and Interpreting Analyses in COMPARE® .....</b>	<b>1</b>
	Introduction.....	1
<b>II.</b>	<b>Installing COMPARE .....</b>	<b>3</b>
	Minimum System Requirements .....	3
<b>III.</b>	<b>Steps for Conducting Compensation Analyses in COMPARE 5</b>	<b>5</b>
	Step 1 – Preparing the Data.....	5
	Step 2 – Opening the Data to be Analyzed .....	7
	Step 3 – Opening the COMPARE Excel Workbook.....	7
	Step 4 – Configuring the COMPARE Data Setup .....	8
	Step 5 – Configuring the COMPARE Analysis Setup .....	9
	Step 6 – Configuring the COMPARE Report Setup .....	11
	Step 7 – Configuring the COMPARE Analysis Specifications .....	12
	Step 8 – Running the Analysis.....	13
<b>IV.</b>	<b>Interpreting Compensation Analysis Results in COMPARE ..</b>	<b>15</b>
	<b>Section 1 – Evaluating the Model .....</b>	<b>16</b>
	Worksheet 1 – Summary:.....	16
	Worksheet 2 – Descriptives: .....	19
	Worksheet 3 – Correlations:.....	21
	Worksheet 4 – Multicollinearity: .....	22
	Worksheet 5 – Diagnostics: .....	23
	Worksheet 6 – Outliers: .....	25
	Worksheet 7 – Interaction: .....	26
	<b>Section 2 – Evaluating the Statistical Results.....</b>	<b>28</b>
	Worksheet 8 – Stats (No Exp Var):.....	28
	Worksheet 9 – Stats(Exp Var):.....	29
	Worksheet 10 – Compensation Analysis: .....	30
	Worksheet 11 – Liabilities: .....	32



# I. Conducting and Interpreting Analyses in COMPARE<sup>®</sup>

---

## Introduction

This manual serves as an introduction and step-by-step guide to conducting compensation analyses using BCG's COMPARE compensation analysis software package. It is strongly recommended that the information presented throughout this book be mastered *prior* to conducting an analysis with COMPARE. The nature of statistical software is that it can make complex analyses *appear* simple. However, given the data assumptions and analytical requirements discussed previously, much can go wrong in the course of a multiple regression analysis. Relevant issues must be identified and addressed to ensure that results are as accurate, valid, and defensible as possible.

---

***WARNING: Multiple Regression is an extremely powerful and highly complex statistical procedure. While it is the ideal tool for determining whether differences in pay are due to "legitimate" factors, or possibly due to protected employee characteristics, if it is not applied correctly it can lead to misleading and/or unsupported results. If the analyst is not completely knowledgeable and experienced with the concepts presented in this manual, or if the employer is considering making changes to employee compensation as a result of the analyses, it is strongly encouraged that assistance be enlisted from an attorney or a statistical professional familiar with the use of multiple regression in compensation analysis.***

---



# II. Installing COMPARE

COMPARE was developed to seamlessly integrate with Microsoft™ Excel™. The program was developed from the ground up as a suite of stand-alone modules assembled together in an Excel workbook. Installation is simply a matter of opening the COMPARE Excel workbook the same way any other Excel file would be opened.

Activation is required after the user has downloaded COMPARE from [www.BCGinstitute.org](http://www.BCGinstitute.org) or installed it from the CD. The program will provide an Activation ID that must be registered for COMPARE to operate.

---

## Minimum System Requirements

Minimum installation requirements include:

- Microsoft Windows XP, Vista, or 7
- Microsoft Excel 2003 or 2007

---

***Important Note: Depending on your security settings, you may receive a warning that COMPARE requires that Excel macros be enabled. Some organizations set their security settings so that macros are disabled by default (to protect against viruses or malicious software). The user can adjust Excel to permit COMPARE to run but not other macros. If you have additional questions, contact your organization's Information Technology (IT) Department or Technical Support at Biddle Consulting Group at 1-800-999-0438.***

---





# III. Conducting Compensation Analyses in COMPARE

There are eight steps to running a complete compensation analysis in COMPARE. The following is a brief overview. Details will be presented in subsequent sections.

1. Prepare the Data. Prepare the data to the specifications detailed in the *Data Preparation* section below.
2. Open the Data to be Analyzed. Before starting COMPARE, the prepared data file must be opened.
3. Open the COMPARE Excel Workbook.
4. COMPARE Data Setup. Specify:
  - Data Workbook
  - Data Worksheet
  - Output Workbook
5. COMPARE Analysis Setup. Specify the compensation model, including:
  - Salary (Required)
  - Explanatory Variable(s) (Optional)
  - Comparison Group Variable (Required)
  - Analyze By Variable(s) (Optional)
  - Employee Identifier (Optional)
6. COMPARE Report Setup (Optional). Customize the reports to be generated and some details in each of the reports (if desired).
7. COMPARE Analysis Specifications (Optional). Customize analysis specifications (if desired).
8. Run the Analysis. COMPARE begins analyzing the data and generating reports.

## **Step 1 – Preparing the Data**

The purpose of data preparation is to ensure that the data to be analyzed conforms to COMPARE's required data specifications.

1. Data Worksheet Setup: To ensure that COMPARE can import the data, the data worksheet must conform to the following basic rules:
  - Data Rows: Variable Names must be on Row 1.
  - Data Columns: Empty columns are not allowed. COMPARE will ignore all data columns to the right of any empty columns.
  - Naming Variables: No spaces are allowed in Variable Names. For example, *GenderCoded (without spacing)* is acceptable, but *Gender Coded (with spacing)* is not.

**Data and Variable Setup:** Compensation analyses have very specific requirements and very specific “data types” for each field included in the analysis. For example, the salary variable can only be numeric (e.g., \$50,000) and the comparison group variable, which identifies the group membership of each record, must be binary (e.g., 0=Women, 1=Men). The following section will begin with a discussion of datatype then move into the specifics of COMPARE’s data requirements.

**Datatype:** COMPARE was designed to import three datatypes as shown below:

Datatype		Details
Character Strings	<STR>	Alphabetical characters and words (e.g., Unit01, Job2)
Numeric	<NUM>	Numbers only (e.g., \$50,000, 3, 17)
Binary	<BIN>	Only 0 and 1

**Data Requirements:** COMPARE only allows certain datatypes for each of the fields on the *Analysis Setup* screen (Step 5). The following table is a quick reference of the variable fields on *Analysis Setup* and the permissible datatype(s).

Variable Fields	Datatype			Required
	Character String <STR>	Numeric <NUM>	Binary <BIN>	
1. Salary		✓		✓
2. Explanatory		✓	✓	
3. Comparison Group			✓	✓
4. Analyze By	✓	✓	✓	
5. Employee Identifier		✓		

1. **Salary:** The salary variable is a required field and must be numeric. As a word of caution, some databases may include characters in the salary variable to track events, and COMPARE will treat that variable as a character strings <STR>. Consequently, the analyst will not be able to specify that variable into the salary field.
2. **Explanatory:** Explanatory variables help to explain potential pay differences (e.g., tenure, performance, years of education), though they are optional. COMPARE will only allow Numeric <NUM> and Binary <BIN> datatypes to be specified in this field. If the analyst wants to include explanatory variables in the compensation model, they must ensure they are Numeric or Binary datatypes.<sup>1</sup>
3. **Comparison Group:** COMPARE tests for pay disparities between two groups—Focal and Reference (e.g., men/women). The comparison group variable identifies the membership of each data record into one of two groups, therefore, it can only be a Binary <BIN> datatype. As part of data preparation, the analyst must code out members of each group as follows:
  - Focal members (e.g., women/minority/individual minority group/40+) = 0
  - Reference members (e.g., men/white/<40) = 1

4. **Analyze By:** If employees can be separated into “Similarly Situated Employee Groupings” (SSEGs), it is important to run compensation analyses for each of these SSEGs. Appropriate SSEGs are *critical* to a valid compensation analysis. Analysis results are invalid and misleading if they are drawn from data with employees who do not perform similar work functions or who are expected to have similar pay, such as if office assistants, research directors, and interns were all clustered into one analysis.  
Simply create a variable (or variables) to identify SSEGs (e.g., job title, region, business unit, AAPs). The SSEG variable can be Character String <STR>, Numbers <NUM>, or Binary <BIN>.
5. **Employee Identifier:** This is an optional specification, but it is helpful to include employee ID in the final reports (e.g., identifying outliers, liabilities-distribution). This field must be Numeric <NUM>.

A sample dataset may look like the following:

Sample COMPARE Dataset

ID	Salary	Tenure	Performance	Gender	GenderCoded
41	34,641	8	1	Man	1
78	34,107	7	1	Woman	0
48	34,550	8	4	Man	1
22	35,832	3	5	Man	1
81	34,884	2	5	Woman	0
90	35,309	3	5	Man	1
92	34,016	3	2	Woman	0
84	34,286	5	2	Woman	0
43	35,809	3	5	Woman	0

**Step 2 – Opening the Data to be Analyzed**

COMPARE requires that the data to be analyzed be opened prior to starting the program. If the data to be analyzed is opened, then it will be available to be selected in the *Select Workbook* drop-down discussed in Step 4, below.

**Step 3 – Opening the COMPARE Excel Workbook**

COMPARE was developed to be seamlessly integrated with Excel. To access COMPARE’s suite of statistical modules, simply open it as any Excel workbook and the following screen should appear.

Click **Run** to begin.



## **Step 4 – Configuring the COMPARE Data Setup**

The first step is to specify the *Input Data Source* and the destination for the output reports.

### **Input Data Source**

Users must specify the Excel Workbook and Worksheet of the data to be analyzed:

1. **Select Workbook:**  
Select the workbook containing the data to be analyzed. Note that the workbook must already be open as instructed in Step 2. To select the workbook, simply click the drop-down arrow next to the *Select Workbook* box. Next, click the **Select** button to instruct COMPARE that the workbook has been selected.
2. **Select Worksheet:**  
Select the worksheet containing the data to be analyzed. Note that the only worksheets available to select are those contained in the workbook previously selected. If there are multiple sheets in the workbook, be sure to select the worksheet containing the data that has been specifically prepared for analysis. To select the worksheet, click the drop-down arrow next to the *Select Worksheet* box. Next, click the **Select** button to instruct COMPARE that the worksheet has been selected.



### **Output Workbook**

The final step of data setup is to select a workbook into which the results of the analysis will be written. Two options exist:

1. **Saving Results to a New Workbook:** To save the results to a new workbook, click the down arrow next to the *Output Workbook* box and click **New**. Then click **Select**. The results will be placed into a new Excel workbook.
2. **Saving Results to an Existing and Opened Workbook:** To save the results into an existing workbook, that workbook must be opened at this point for it to be listed in the drop-down selection box. Click the down arrow next to the *Output Workbook* box and select the name of the workbook into which to place the analysis results. Then click **Select**.

Once the *Data Input Source* and *Output Workbooks* have been selected, click the **Next** button.

## Step 5 – Configuring the COMPARE Analysis Setup

This screen is where the analyst specifies the compensation model (i.e., the variables that belong in the compensation analysis).

### Analysis Setup

There are six major fields on this screen:

1. Available Variables
2. Salary
3. Explanatory Variables
4. Comparison Group Variable
5. Analyze By Variable(s)
6. Employee Identifier

To specify a variable field, simply click the arrow buttons   associated with each of the variable fields. To add a variable to a field, click the right-arrow button . To remove a variable from a field, select the variable and click the left-arrow button .

Compare 2.x -- Biddle Consulting Group, Inc. Copyright (2009)

Data Setup | Analysis Setup | Report Setup | Analysis Specifications

Available Variables

- <NUM> ID
- <STR> Birthdate
- <NUM> Educ
- <STR> Bunit
- <NUM> JobGroup
- <NUM> Salary
- <NUM> SalaryStart
- <NUM> PPA
- <NUM> PPA22
- <NUM> TIC
- <NUM> TIJ
- <BIN> Minority
- <BIN> Gender

Salary:

Explanatory Variables:

Comparison Group Variable:

Analyze By Variable(s):

Employee Identifier (Optional):

Analyses

- Descriptives
- Diagnostics
- Correlations
- Interactions
- MultiCollinearity
- Regression

< Previous    Reset    Setup    RUN    Help    END


The six fields are discussed in detail below:


1. Available Variables: The box on the left side of the screen presents a list of the fields that were successfully imported into COMPARE. These variables are available for the compensation analysis.
2. Salary (Required): From the *Available Variables* box, select the salary/compensation variable by clicking on the variable and click the right-arrow button  next to the *Salary* box. Note that only the Numeric <NUM> datatype is permissible.
3. Explanatory Variables (Optional): If there are variable(s) that account for potential pay differences, the analyst may specify them in the *Explanatory Variables* box. For each explanatory variable, select it from the *Available Variables* box and click the right-arrow button  next to the *Explanatory Variables* box. Note that only the Numeric <NUM> datatype is permissible. The following explanatory variables will be used in examples throughout this manual:
  - TIC = Time in Company
  - TIJ = Time in Job
  - PPA = Performance Appraisal Score(s)
4. Comparison Group Variable (Required): This is one of two required specifications and is critical to specifying the correct compensation analysis. This variable identifies members of groups (0=Focal, 1=Reference) to compare in the compensation model. From the *Available Variables* box, select the grouping variable and click the right-arrow button  next to the *Comparison Group Variable* box. Note that only the Binary <BIN> datatype is permissible.

5. Analyze By Variable(s) (Optional): Specifying the *Analyze By* variable is optional, but it should not be ignored if employees in the data are grouped into “Similarly Situated Employee Groupings” (SSEGs).<sup>2</sup>

COMPARE can easily run multiple analyses simultaneously (e.g., by department + job title, etc.). Currently, the program can handle up to five strata. However, based on previous experience, there is rarely ever a need to overlap more than three strata. For example, a typical three strata specification would be:

AAP/Location x Job Group x JobTitle

To specify the analysis strata, simply select each SSEG-layer variable and click the right-arrow button  next to the *Analyze By Variable(s)* box. It is important to note that the order in which variables are placed into this box will determine the way in which the results are displayed. For example, if Location is selected first, followed by Job Group, the results will display all Job Groups within each Location. However, if Job Group is selected first, followed by Location, then the results will display all Locations within each Job Group.

Employee Identifier (Optional): Unique employee IDs are helpful references when interpreting the Outliers and Liability computation reports. If an Employee Identifier is not specified, COMPARE will generate a unique and random Employee ID for each record. To specify an employee identifier, select the variable from the *Available Variables* box, and click the right-arrow button  next to the *Employee Identifier* box. Note that only the Numeric <NUM> datatype is permissible in this specification.

### Analyses

By default, COMPARE will run a comprehensive set of model diagnostics and statistical tests for analyzing pay disparities, including *Descriptives*, *Diagnostics*, *Correlations*, *Interactions*, *Multicollinearity*, and *Regression*. Of the six analyses, *Descriptives* and *Regression* are required for each COMPARE analysis. The analyst may choose not to run the remaining four analyses.

### Setup: Advanced User Specifications

COMPARE has a very efficient user interface requiring minimal user specifications. It should be recognized, however, that highly-complex statistical procedures are run from this relatively simple interface.

Analysts may choose to rely on default settings or specify additional analyses to better suit their statistical needs. The default advanced-user specifications have been set to BCG best practice-standards and are suitable for most compensation analyses.

Advanced users may adjust these settings by clicking  to access two power-user specification screens: 1) Report Setup and 2) Analysis Specifications. Each of these will be detailed in the next section.

## Step 6 – Configuring the COMPARE Report Setup

The *Report Setup* screen allows the analyst to customize the output reports. There are five major specifications that the analyst can customize: Output Reports, Outlier Analysis, Group Heading Specifications, Data Worksheets, and Report Header.

### Report Setup

1. **Output Reports:** By default, COMPARE generates a comprehensive library of reports. However, there are three reports that are not automatically generated. These are all raw statistical outputs from the regression analysis, *t*-Test analysis, and the liabilities analysis.

Analysts who wish to evaluate these detailed reports can select them from the *Output Reports* options.

2. **Outlier Analysis:** COMPARE identifies outliers by computing their standardized *z*-values to determine how far a value may be from the mean. By default, COMPARE only flags outliers that exceed 1.96 standard deviation units without specifying the exact *z*-value.

Analysts who wish to evaluate the actual *z*-scores may obtain them by selecting the *Outlier Analysis* option.

3. **Group Heading Specifications:** By default, COMPARE labels the group headings as either Focal or Reference.

Analysts may choose to customize their reports by entering more specific labels based upon the analyses being conducted.

4. **Data Worksheets:** In the output reports workbook, COMPARE does not automatically include the analysis dataset or copies of the original data.

If the analyst desires to replicate the analysis data in the output workbook, it is possible to request copies of the original data:

- *Analysis Data* is the subset of variables and values specified in Step 5 (*Analysis Setup*).
- *Copy of Original* is a copy of the original dataset in its entirety.

5. **Report Header:** COMPARE includes an Attorney-Client Privileged header across all reports. The analyst may change this header by editing the field, though it is recommended to leave it in place.

Compare 2.x -- Biddle Consulting Group, Inc. Copyright (2009)

Data Setup | Analysis Setup | Report Setup | Analysis Specifications

**Output Reports**

Descriptive     Full Regression     Full Liability

Summary     Full t-test

**Outliers Analysis**

Outlier z-Scores

**Group Heading Specifications**

Focal Group:   
(e.g., Female, Minority, 40+)

Reference Group:   
(e.g., Male, Non-Minority, <40)

**Data Worksheets**

Analysis Data

Copy of Original

**Report Header**

Help    Reset    Done



## Step 7 – Configuring the COMPARE Analysis Specifications

On the *Analysis Specifications* screen, the user can customize detailed analysis settings. There are three major specification groups: 1) Liabilities Analysis, 2) Warning Thresholds, and 3) Sample Size Requirements.

### Analysis Specifications

1. Liabilities Analysis: The following are the default specifications:

- Compute Liabilities: COMPARE computes liabilities by default. If the analyst does not wish to generate Liabilities, select *No*.
- Standard Deviation Diff: By default, liabilities are computed to reduce the pay disparity to one standard-deviation difference. For the advanced analyst, it is possible to modify this by selecting from the drop-down menu or entering a specific value.

- Compute Reference Group Liabilities: Compensation analyses evaluate impact against both Focal and Reference group members. This is particularly relevant as Title VII enforcement agencies pursue and successfully bring claims where the impact has been against reference group members. Consequently, COMPARE computes liabilities for both Focal and Reference group members by default. The analyst who wishes only to compute liabilities for Focal group members may choose to uncheck this selection.
- Liabilities Distribution: There are three methods of distributing liabilities:
  - Evenly Distributed: All Individuals. Evenly distributes liabilities to all individuals in the negatively-impacted group.
  - Evenly Distributed: Below Average. COMPARE identifies individuals in the impacted group who are below the group average and evenly distributes the liability to them.
  - Proportionally Weighted by Impact. COMPARE identifies all individuals in the negatively-impacted group and computes the difference between their current salary and their predicted salary (based on a model using only job-related qualifications and *not* gender/minority status). Only individuals in the negatively-impacted group who fall below their predicted compensation are identified as eligible to receive a portion of calculated liability amount. Among those individuals, liability amounts are computed in proportion to the difference between the individuals' current salary and their predicted salary (see *Compensation Analyses*:

*A Practitioner's Guide to Identifying and Addressing Compensation Disparities* by Biddle Consulting Group for a full description of this method).

2. Warning Threshold Specifications: COMPARE has built-in features to highlight and flag areas of potential concern. The default settings for triggering these warnings are based on best practices and statistical standards. Advanced analysts may change these thresholds to better suit their needs. For most cases, however, it is advisable not to change the defaults, especially if the analyst has no clear justification to change the sensitivity of these tests.

Potential Liability: COMPARE provides a very powerful and sensitive test to help identify potential liabilities. It is descriptive-based rather than statistical, but is very helpful in identifying areas that may trigger further investigations from federal enforcement agencies. Groups flagged with a “medium” or “high” potential liability should be given the most attention.

COMPARE flags three levels of potential disparity (Low, Medium, and High) by evaluating three descriptive factors:

- All Total (#): This is the total number of individuals in each SSEG. The larger the number of employees, the higher the likelihood of potential liability if there are differences in average compensation. The default labeling is Low=10, Medium=20, High=30.
- Subgroup (#): The number of individuals in each subgroup (e.g., men/women) within each SSEG. Even if there is a large number of total employees within the SSEG, if there are not many in both groups then statistical analyses become difficult and the potential liability is decreased. The default labeling is Low=3, Medium=6, High=10.
- Group Diff. (%): This is the difference in average salary between Focal and Reference groups. The larger the difference the greater the potential liability. The default is to flag groups with differences greater than 5%.

Multicollinearity: COMPARE provides two measures of multicollinearity: VIF and Pearson’s  $r$  (to identify correlations between variables). The default thresholds reflect well-established statistical standards.

Outlier: COMPARE identifies outliers by computing the  $z$ -score for each value by variable. The default threshold of  $z > 1.96$  will flag values that are statistically significantly different than the mean (i.e., they exceeds the 95% confidence interval around the mean).

3. Sample Size Requirements: For statistical tests to provide valid conclusions, the sample sizes need to be sufficient. COMPARE evaluates sample size on two levels:
  - All Total (#): Total number of records (i.e., employees) in each SSEG.
  - Subgroup(#): The number of employees in each subgroup (e.g., men/women) within an SSEG.

SSEGs that do not meet these very low sample size requirements for an analysis will not be analyzed.

After specifying the *Report Setup* and *Analysis Specifications*, return to the *Analysis Setup* screen to run the analysis by clicking on the  button.

---

***Important Note: Changes to the default settings within COMPARE are not saved. If the default settings are changed, they will need to be changed again for each subsequent set of analyses.***

---

## **Step 8 – Running the Analysis**

Once the compensation model has been specified (i.e., variables moved to the appropriate boxes) and the reports have been selected and configured, the next step is for COMPARE to run the battery of analyses and generate statistics that will help evaluate whether the model is statistically significant, whether there are unexplained differences in pay by group (e.g., gender) after controlling for legitimate job-related explanatory factors, and where pay adjustments may be necessary.

To run the analysis, simply click **Run** on the *Analysis Setup* screen. If the data set is large or if there are many different strata (i.e., multiple split levels), the analysis may take some time to complete. The *Analysis Status* window should indicate the progress of the analysis. If a problem occurs, please notify COMPARE technical support (support@biddle.com).

### Analysis Summary and User Warning

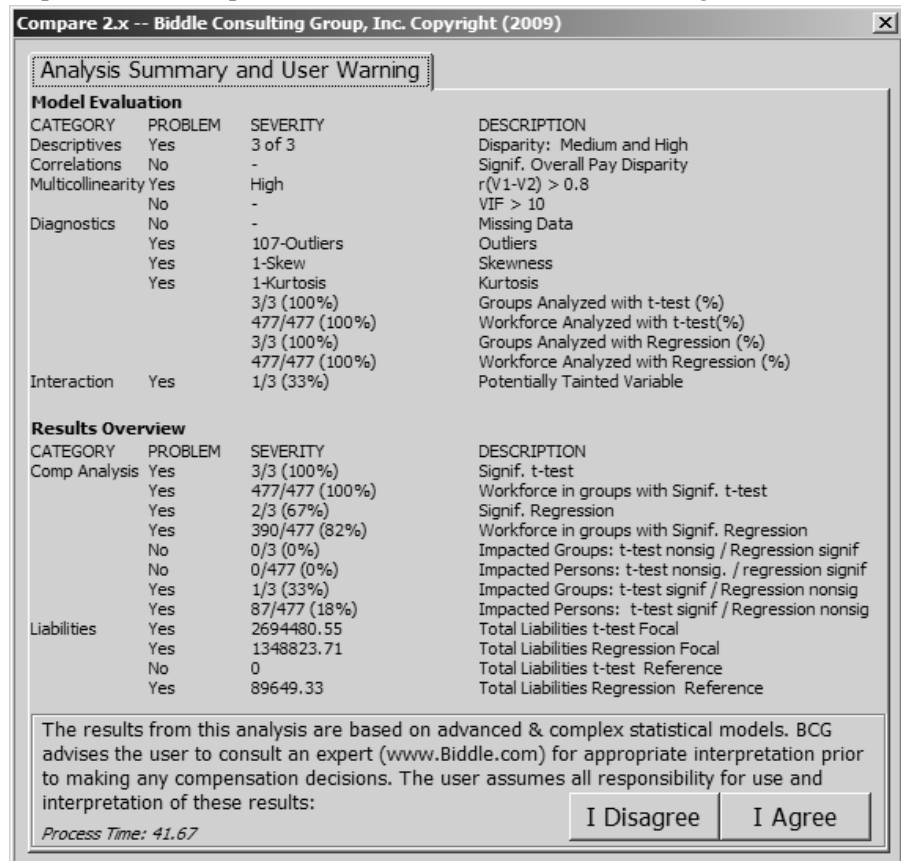
Once COMPARE completes all computations and reports, the results are summarized in a high-level overview.

The *Analysis Summary and User Warning* screen has four major columns:

1. **Category:** Identifies the category of analysis performed.
2. **Problem:** Identifies whether a problem exists by analysis category.
3. **Severity:** Identifies the severity of the problem by analysis category.
4. **Description:** Provides a brief description of the problem and some high-level details.

The summary and all reports reflect the principle that the validity of the compensation model should be evaluated first and the results second. The *Analysis Summary and User Warning* screen has two major sections: *Model Evaluation* and *Results Overview*.

1. **Model Evaluation:** Lists any problems with the data identified during diagnostics that may undermine the validity of the statistical analyses. The model evaluation summary includes warnings for Potential Disparity, Significant Correlations, Multicollinearity, Data Diagnostics (e.g., outliers, skewness, kurtosis), Proportion of Data Analyzed, and Interactions.
2. **Results Overview:** Lists a summary of significant and non-significant *t*-Tests and regression analyses, and total liabilities by statistical model.




---

**Important Note:** *It is critical that this screen be carefully reviewed. COMPARE was designed with an efficient and user-friendly interface with the intent to allow non-statistical analysts to run very complex statistical analyses. However, unless the data is accurate, assumptions are met, and the results have been validated, no conclusions can be made with confidence.*

---

On this screen, COMPARE requires analysts to acknowledge that, given the complexity of multiple regression, they are responsible for ensuring they understand the findings and that Biddle Consulting Group, Inc. and its staff will not be held liable for any use of these results.

Click on **Agree** to receive access to the full analysis report.

# IV. Interpreting Compensation Analysis Results in COMPARE

Once COMPARE completes all analyses and the user has agreed with the warning screen and caveats, the following Excel worksheets will be presented. These worksheets provide all of the information necessary to evaluate compensation disparities.

The default reports can be segmented into two major sections and will contain the following individual worksheets:

---

Section 1: Model Evaluation	Section 2: Analysis Results
Worksheet 1: Summary	Worksheet 8: Stats(No Exp Var)
Worksheet 2: Descriptives	Worksheet 9: Stats(Exp Var)
Worksheet 3: Correlations	Worksheet 10: Comp Analysis
Worksheet 4: Multicollinearity	Worksheet 11: Liabilities
Worksheet 5: Diagnostics	
Worksheet 6: Outliers	
Worksheet 7: Interaction	

---

As a matter of practice, the analyst should review the worksheets in sequential order, from left to right. The reports in each worksheet are ordered to facilitate proper regression model evaluation prior to reviewing actual results of the analyses. The information in each tab should be carefully reviewed and clearly understood before moving to the next tab. Each of these tabs will be discussed below.

---

***WARNING: Carefully review and understand each tab in the output workbook, from left to right, before moving to the next. Review all tabs before attempting to interpret the analysis results.***

***Interpretation of, and descriptions for, each of the report worksheets are presented in order in the following sections.***

---

# Section 1 – Evaluating the Model

## Worksheet 1 – Summary:

The *Summary* worksheet is the core of the program because it quickly presents the analyst with warnings and a summary regarding whether problem areas were found with employee compensation. There are two primary sections to this report: *Model Evaluation* and *Results Summary*. The components of this worksheet are presented below.

### Model Evaluations

Category	Problem	Severity	Description
Descriptives	Yes	3 of 3	Disparity: Medium and High
Correlations	No	-	Signif. Overall Pay Disparity
Multicollinearity	Yes	High	$r(V1-V2) > 0.8$
	No	-	VIF > 10
Diagnostics	No	-	Missing Data
	Yes	107-Outliers	Outliers
	Yes	1-Skew	Skewness
	Yes	1-Kurtosis	Kurtosis
		3/3 (100%)	Groups Analyzed with <i>t</i> -Test (%)
		477/477 (100%)	Workforce Analyzed with <i>t</i> -Test (%)
		3/3 (100%)	Groups Analyzed with Regression(%)
	477/477 (100%)	Workforce Analyzed with Regression (%)	
Interaction	Yes	1/3 (33%)	Potentially Tainted Variable

- **Descriptives:** “YES” indicates COMPARE identified one or more areas of medium or high potential liability. These differences are based on descriptive statistics, but are very helpful at identifying potential problem areas.
- **Correlations:** “YES” indicates a statistically significant correlation between the comparison group membership (e.g., whites or minorities) and compensation. A significant relationship between these two variables is a potential indicator of discrimination.
- **Multicollinearity:** “YES” indicates the existence of multicollinearity and the model’s findings may be suspect.
- **Diagnostics:** “YES” indicates a problem with the indicated diagnostic as well as the number of employee groups which were analyzed with either a *t*-Test or multiple regression.
- **Interaction:** “YES” indicates one or more variables were interacting with the comparison group and may be potentially tainted.

## Results Summary

Category	Problem	Severity	Description
Comp Analysis	Yes	3 of 3 (100%)	Signif. <i>t</i> -Test
	Yes	477/477 (100%)	Workforce in Groups with Signif. <i>t</i> -Test Difference
	Yes	2/3 (67%)	Signif. Regression
	Yes	390/477 (82%)	Workforce in Groups with Signif. Regression Difference
	No	0/3 (0%)	Non-Signif. <i>t</i> -Test but Signif. Regression
	No	0/477 (0%)	Workforce in Groups with Non-Signif. <i>t</i> -Test but Signif. Regression Difference
	Yes	1/3 (33%)	Signif. <i>t</i> -Test but Non-Signif. Regression
	Yes	87/477 (18%)	Workforce in Groups with Non-Signif. Regression but Signif. <i>t</i> -Test Difference
Liabilities	Yes	\$2,694,480.55	Total Liabilities <i>t</i> -Test (Focal)
	Yes	\$1,348,823.71	Total Liabilities Regression (Focal)
	No	\$0.00	Total Liabilities <i>t</i> -Test (Reference)
	Yes	\$89,649.33	Total Liabilities Regression (Reference)
	Yes	\$11,027.23	Liability Excluded: N=10

- **Comp Analysis:** This section of the report provides several high-level perspectives on compensation differences.
  - **Signif. *t*-Test:** Provides a count and percentage of SSEGs (e.g., job groups, job titles, etc.) that exhibited significant pay disparities based upon *t*-Test results.
  - **Workforce in Groups with Signif. *t*-Test Difference:** Provides a count and percentage of the total workforce in SSEGs with significant pay disparities based upon *t*-Test results.
  - **Signif. Regression:** Provides a count and percentage of the SSEGs that exhibited significant pay disparities based upon multiple regression results.
  - **Workforce in Groups with Signif. Regression Difference:** Provides a count and percentage of the total workforce in SSEGs with significant pay-disparities based upon multiple regression results.
  - **Non-Signif. *t*-Test but Signif. Regression:** Provides a count and percentage of the number of SSEGs that were not-significant based upon *t*-Test results but became significant when using multiple regression.
  - **Workforce in Groups with Non-Signif. *t*-Test but Signif. Regression Difference:** Provides a count and percentage of the workforce in SSEGs where pay disparities were not significant when using a *t*-Test but became significant when using multiple regression.

- Signif. *t*-Test but Non-Signif. Regression: Provides a count and percentage of the number of SSEGs that were significant based upon *t*-Test results but became non-significant when using multiple regression.
- Workforce in Groups with Signif. *t*-Test but Non-Signif. Regression: Provides a count and percentage of the workforce in SSEGs where pay disparities were significant when based upon *t*-Test results but became non-significant when using multiple regression.
- Liabilities: This category provides several high-level summary of liability by statistical model (*t*-Test/regression) and group membership (Focal/Reference).
  - Total Liabilities *t*-Test (Focal): Provides the total dollar amount for Focal group members (e.g., women/minorities, etc.) necessary to eliminate the statistically significant differences in compensation summed across all analyses (based on *t*-Test results).
  - Total Liabilities Regression (Focal): Provides the total dollar amount for Focal group members (e.g., women/minorities, etc.) necessary to eliminate the statistically significant differences in compensation summed across all analyses (based on multiple regression results).
  - Total Liabilities *t*-Test (Reference): Provides the total dollar amount for Reference group members (e.g., women/minorities, etc.) necessary to eliminate the statistically significant differences in compensation summed across all analyses (based on *t*-Test results).
  - Total Liabilities Regression (Reference): Provides the total dollar amount for Reference group members (e.g., women/minorities, etc.) necessary to eliminate the statistically significant differences in compensation summed across all analyses (based on multiple regression results).
  - Liability Excluded: Provides the number of focal group members that are excluded from the regression (and thus the liabilities) calculated by COMPARE. For example, if 50 women are included in the analysis and two explanatory variables are included in the model (e.g., tenure and job performance scores), but complete data is only available for 40 of the women, the program will display the error noted in the Results Summary above. When this error appears, Full Information Maximum Likelihood (FIML) statistical modeling should be used to impute data for the missing records. Call BCG for assistance.

---

***Important Note: It is strongly recommended that no adjustments in pay be made without first validating the model and exploring whether other factors that were not included in the model but which are otherwise legitimate may be influencing pay. It is recommended that analysts and employers seek qualified statistical assistance at this stage.***

---

## **Worksheet 2 – Descriptives:**

The *Descriptives* worksheet presents descriptive statistics for each of the variables included in the analysis broken down by each level of the *Analyze by Variable* (SSEGs) selected during *Analysis Setup*. The components of this worksheet are presented below:

### ***Descriptives Report***

Group	Data	Focal	Reference	Grand Total	Impacted/ Difference	Diff. (%)	Potential Liability
1	Count of Minority	276	87	363	Focal	17.16	High
	Avg of Salary	\$18,414.38	\$22,230.11	\$19,328.90	\$3,815.73		
	Avg of PPA	2.0	1.8	1.9			
	Avg of TIC	4.7	17.1	7.7			
	Avg of TIJ	4.6	16.9	7.5			
2	Count of Minority	14	13	27	Focal	6.51	Medium
	Avg of Salary	\$28,211.36	\$30,175.77	\$29,157.19	\$1,964.41		
	Avg of PPA	3.2	3.0	3.1			
	Avg of TIC	11.1	28.2	19.4			
	Avg of TIJ	2.9	29.2	15.5			
3	Count of Minority	80	7	87	Focal	33.61	Medium
	Avg of Salary	\$49,314.03	\$74,279.14	\$51,322.71	\$24,965.12		
	Avg of PPA	4.2	5.7	4.4			
	Avg of TIC	6.8	28.1	8.6			
	Avg of TIJ	5.2	16.3	6.1			

- **Analyze By/SSEG:** Represents the analysis groups selected by the analyst (e.g., Job Group).
- **Data:** Indicates the type of descriptive statistics computed.
- **Focal/Reference:** Provide descriptive statistics for Focal and Reference group members.
- **Grand Total:** Provides summary descriptive statistics for the entire group (Focal+Reference).
- **Impacted/Difference:** Provides two pieces of high-level information:
  - Indicates which group is negatively impacted.
  - Provides the raw mean difference between Focal and Reference group members.
- **Difference (%):** Provides the percent difference in compensation between Focal and Reference group members. The higher paid group is used as the computed reference.



- Potential Liability:<sup>3</sup> Indicates whether the potential for liability is Low, Medium, or High based on counts of Focal and Reference members and percent difference in compensation. High potential liability areas are shaded.

### **Worksheet 3 – Correlations:**

The *Correlations* worksheet is particularly useful for helping identify those factors that are most powerful in explaining the company's pay practices. The components of this worksheet are presented below:

#### Correlation Report

Variable Name	Salary	PPA	TIC	TIJ	Minority
Salary	--	0.849**	0.217**	0.062	0.034
PPA	0.849**	--	0.1*	-0.042	-0.091*
TIC	0.217**	0.1*	--	0.902**	0.813**
TIJ	0.062	-0.042	0.902**	--	0.856**
Minority	0.034	-0.091*	0.813**	0.856**	--

Notes:

\* Significant at the .05 level

\*\* Significant at the .01 level

The correlation coefficients included in the *Correlations* worksheet allow the analyst to determine whether:

- Explanatory variables are related to compensation;
- The focal variable (i.e., gender, minority status, etc.) is related to compensation; or
- There are relationships between the various predictors that could lead to multicollinearity.

When reviewing the *Correlation Report*, pay particular attention to the second column (to the immediate right of the *Variable List* column). This column indicates the correlations between each factor and the variable being used to measure employee compensation. Ideally, those variables that are most highly correlated with compensation are the explanatory (i.e., job related) variables and not the protected characteristic (e.g., gender/race). Those factors that share a statistically significant relationship at  $p < 0.05$  are identified with a single asterisk (\*) and those that have a statistically significant relationship at the  $p < 0.01$  level are identified with a double asterisk (\*\*).

Next, examine the correlations between each of the other factors. If a correlation is very high (i.e., approaching 1.00), it indicates a potential problem with collinearity which would mean that both variables cannot be included in the model without leading to serious problems with the regression model. COMPARE includes specific measures of multicollinearity in a separate tab, but this is a useful visual tool as well. If two explanatory factors appear to be too highly related (e.g.,  $r > .80$ ), it may be best to remove the variable of the two with the weakest relationship to compensation

## **Worksheet 4 – Multicollinearity:**

The *Multicollinearity* worksheet helps to identify signs of excessive relationships between explanatory variables. The components of this worksheet are presented below:

### Multicollinearity Report

Var1 (V1)	Var2 (V2)	Correlations			VIF	Multicollinearity? (Y/N)
		V1-V2	DV-V1	DV-V1.V2		
PPA	TIC	0.100	0.849	0.851	1.116	
PPA	TIJ	-0.042	0.849	0.854		
TIC	PPA	0.100	0.217	0.250	5.978	
TIC	TIJ	0.902	0.217	0.373		
TIJ	PPA	-0.042	0.062	0.184	5.929	
TIJ	TIC	0.902	0.062	-0.317		

The following guidelines are adopted by COMPARE:

- **Variance Inflation Factor (VIF):** This is a measure of the impact of collinearity among the variables in a regression model. It is always a number that is equal to or greater than 1.00. Larger numbers indicate greater impact of collinearity. COMPARE uses a default criterion of 10. SPSS provides conservative guidelines that Condition Index values over 15 indicate *possible* problems with collinearity, while values over 30 suggest a *serious* problem with collinearity.
- **Correlation Between Factors:** A second diagnostic of collinearity is based on the size of the correlation between factors. If the correlation is greater than or equal to 0.80, it is a sign that the two variables are too highly related to both be included in the model. In these situations it is recommended that the variable with the strongest relationship to compensation be retained.
- **Multicollinearity Determination:** If VIF is greater than 10.0 or the correlation is greater than or equal to 0.80, multicollinearity is likely to exist if both of the factors remain in the model.

---

***Important Note: If multicollinearity exists in a model, it means that there is an issue with the data that will result in problems and in faulty results. If multicollinearity exists, it must be evaluated before trusting the results of any analysis. If the analyst does not fully understand the issues and methods for addressing multicollinearity, it is strongly recommended that a professional compensation statistician be consulted.***

---

## **Worksheet 5 – Diagnostics:**

Regression and *t*-Tests require that the data meet certain assumptions (e.g., normality, no extremely outliers). COMPARE evaluates the data for any potential violations in assumptions and provides those results in the *Diagnostics* worksheet. The components of this worksheet are presented below:

### Diagnostics Report

JobGroup	Variable	Valid N	Non Missing (%)	Outliers (#)	Skewness	Kurtosis
1	Salary	363	100%	15		
	PPA	363	100%	13		
	TIC	363	100%	24		
	TIJ	363	100%	30		
2	Salary	27	100%	2		Kurt (-)
	PPA	27	100%	1		
	TIC	27	100%	0	Skew (-)	
	TIJ	27	100%	0		
3	Salary	87	100%	5		
	PPA	87	100%	5		
	TIC	87	100%	7		
	TIJ	87	100%	5		

The analyst should carefully review this report to ensure that data meets the necessary assumptions. In particular, note the following columns:

- **Valid N:** This column indicates the number of valid cases associated with each variable. For example, if a job title has 20 employees but the performance score is only available for 19 of them, the number shown will be 19 rather than 20.
- **Non Missing (%):** This column indicates the percent of the employees for whom valid data is available for all included variables. In the example above, it can be seen that 100% of the data was available for each employee, for each explanatory factor, in the three JobGroups.
- **Outliers (#):** This column allows an analyst to identify the number of outliers that are contained in a particular SSEG. This is important to note as outliers indicate employees who have very high or low values on a variable. These employees can have a dramatic impact on the results of an analysis and can lead to misleading findings of compensation disparities (or the lack thereof).
- **Skewness:** Skewness is caused when outliers tend to exist in only one direction (either the outlier(s) are all high or are all low). COMPARE computes a skew index and determines whether it is significant or not. Only significantly skewed variables are flagged.
  - **Skew (+):** Indicates significant positive skewness. If positive skewness exists, it means there is a group of employees with higher values on a variable than the typical group of employees. Skewed data can lead to misleading results.
  - **Skew (-):** Indicates significant negative skewness. If negative skewness exists, it means there is a group of employees with lower values on a variable than the typical group of employees. Skewed data can lead to misleading results.
  - **Small N:** Indicates that the number of employees in the group is too small for a valid calculation of skewness.
- **Kurtosis:** Kurtosis is a measure of variation which indicates the extent to which employees are spread out on a variable. COMPARE computes a Kurtosis index and determines whether it is significant or not. Kurtosis is flagged only if it is significant.

- Kurt (+): Indicates significant positive Kurtosis, which is a tight-peaked distribution. Kurtosis in the data may mean a violation of normality required by multiple regression analyses.
- Kurt (-): Indicates significant negative Kurtosis, which is a wide-low distribution. Kurtosis in the data may mean a violation of normality required by multiple regression analyses.
- Small N: Indicates that the number of employees in the group is too small to allow for a valid calculation of kurtosis.

## **Worksheet 6 – Outliers:**

The *Outliers* worksheet provides a detailed look at specific employees identified in the *Diagnostics* worksheet. Those identified as having values on a variable that are either unusually high or low compared to other employees are highlighted. The components of this worksheet are presented below:

### Outliers Report

JobGroup	Emp ID	Salary	PPA	TIC	TIJ	Minority
1	376	\$17,416	2.0	8.0	5.0	0
	28	\$17,179	2.0	7.0	5.0	0
	43	\$15,832	1.0	1.0	3.0	0
	36	\$17,546	2.0	10.0	8.0	0
	369	\$16,713	2.0	4.0	2.0	0
	56	\$17,055	2.0	2.0	3.0	0
	90	\$14,286	1.0	3.0	5.0	0
	91	\$15,809	1.0	0.0	3.0	0
	472	\$25,821	3.0	7.0	6.0	0
	5	\$29,380	4.0	7.0	8.0	0
	17	\$30,358	4.0	2.0	1.0	0
	52	\$29,591	4.0	4.0	1.0	0
	80	\$30,485	4.0	0.0	3.0	0
	234	\$29,435	4.0	3.0	1.0	0
	310	\$30,122	4.0	5.0	8.0	0
	319	\$28,010	4.0	-1.0	1.0	0
	72	\$30,415	5.0	3.0	4.0	0

Without clear justification, it is recommended that the analyst not attempt to remove outliers from the data because it could be seen as manipulating the data in a way that gives the results a company desires. Rather, each outlier should be carefully examined to determine whether there are legitimate reasons for their scores to be so different from the typical employee. This information will be useful when evaluating whether all employees in the SSEG are similarly situated (outliers will be flagged in red) or in the event a cohort analysis becomes necessary as a result of statistically significant multiple regression findings.

## **Worksheet 7 – Interaction:**

A statistical interaction occurs when an explanatory variable works differently for one group versus another. These relationships are investigated in the *Interactions* worksheet. The components of this worksheet are presented below:

### Interaction Report

Grade	Job Code	Interaction	R2-Change	<i>r</i> (partial)	<i>t</i> -Value	<i>p</i> -Value
1	Job ABC	Minority x PPA	0.014	0.226	4.533	0.000
		Minority x TIC	0.001	-0.033	-0.654	0.514
		Minority x TIJ	0.003	-0.059	-1.159	0.247

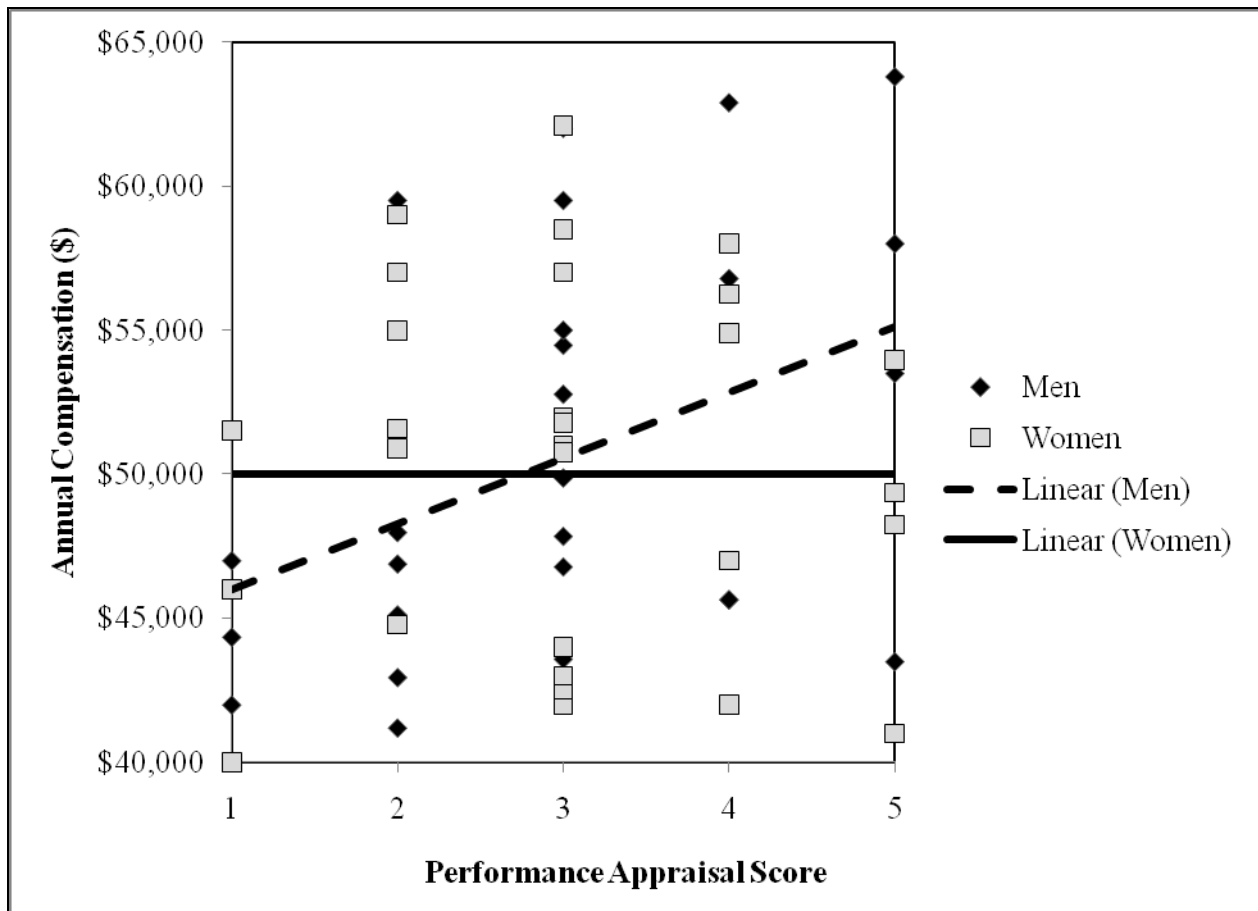
The primary role for COMPARE is to simplify the process of conducting a valid multiple regression analysis for those who may not be professional statisticians but who are required to use this statistical technique to evaluate compensation. Specifically, COMPARE tests for interactions between the *Test Group Variable* (i.e., Focal/Reference, gender, minority status) and each specified *Explanatory Variable*, by SSEG. The user can either run one overall interaction report for the entire dataset or individual interactions for each SSEG. COMPARE summarizes the results in the *Interaction* worksheet.<sup>4</sup>

There are at least five columns on the *Interaction* report:

- 1-2. (Optional) Group by: Identifies the grouping variables (e.g., job group, job title, etc.) by which the compensation analyses will be run.
3. Interaction: Identifies the interaction that is evaluated.
4. R2-Change: Identifies the change in the multiple correlation ( $\Delta R^2$ ), which is the incremental variance accounted for by the interaction term. This is valuable information for the advanced analyst who seeks to evaluate the severity of the interaction.
5. *r* (partial): Provides the partial  $R^2$  of the interaction term as unique incremental variance. This is valuable information for the advanced analyst who seeks to evaluate the severity of the interaction.
6. *t*-value: Provides *t*-value for the statistical test of significance.
7. *p*-value: Provides the *p*-value for the obtained *t*-value. Values in this column less than or equal to 0.05 indicate a significant interaction. Significant interactions are highlighted in red.

Once a significant interaction is identified, the analyst should investigate the variable that is significantly interacting with the *Test Group Variable* (e.g., gender/race, etc.) and compensation. This is a complex topic. A thorough discussion of the topic is beyond the scope of this manual. With that said, interpretation often requires a graph of the interaction. See below for an example of an interaction.

## Interaction Between Annual Compensation and Performance Appraisal Score



Note: Scatterplots are not currently included within COMPARE.

In this example, COMPARE found the gender group by Performance Appraisal Score interaction to be significant. Salary increased with higher Performance Appraisal Scores for men, but remained constant across the range of Performance Appraisal Scores for women. Clearly, Performance Appraisal Scores operate differently for men than women when it comes to making pay predictions.

---

***Important Note: This is a minor treatment of a very complex topic. In general, whenever an interaction is present, please consult professional statisticians who have experience with such matters. Interactions can be problematic on three levels. First, it may be an indicator of a tainted variable. Second, this could be interpreted by some as prima facie evidence of discrimination. Third, it is extremely difficult to identify the dollar amounts needed to rectify the situation and to whom those dollars should be allocated. The mechanics of testing for interactions is fairly straightforward, but proper interpretation and understanding of interactions requires a firm grasp of multivariate statistics. Consequently, it is impossible to fully test an interaction and provide simple palatable interpretation. This is full disclosure and warning for the beginning to moderately advanced analysts. Proper evaluation of interactions and their impact on employee compensation is critical to proper compensation analysis.***

---



## Section 2 – Evaluating the Statistical Results

### Worksheet 8 – Stats (No Exp Var):

The raw *t*-Test statistics are presented in the *Stats(No Exp Var)* (“No Explanatory Variables”) worksheet. Because *t*-Tests are simply unique regression models with no explanatory variables, the results are presented in a general regression framework, which provides more diagnostic information.

Proper interpretation of the report requires a firm understanding of regression statistics. The report presents the statistics in two sections: *Overall Model* and *Individual Factors*. The components of this worksheet are presented below:

#### Analysis Statistics: t-Tests

	Overall Model		Individual Factors		
	Statistics	Value	Statistics	Minority	Int( <i>a</i> )
1	$R^2$	0.09	<i>b</i>	3815.73	18414.38
	Sig( <i>p</i> )	0.00	<i>t</i> -value	5.87	57.85
	Valid n	363.00	Sig( <i>p</i> )	0.00	0.00
2	$R^2$	0.18	<i>b</i>	1964.41	28211.36
	Sig( <i>p</i> )	0.03	<i>t</i> -value	2.34	48.44
	Valid n	27.00	Sig( <i>p</i> )	0.03	0.00
3	$R^2$	0.56	<i>b</i>	24965.12	49314.03
	Sig( <i>p</i> )	0.00	<i>t</i> -value	10.48	72.99
	Valid n	87.00	Sig( <i>p</i> )	0.00	0.00

1. **Overall Model:** Interpretation of the *Overall Model* statistics provides insight into the strength of the specified model.
  - **R-Sqr ( $R^2$ ):** Provides an estimate of effect size (i.e., the amount of variance accounted for by the model).
  - **Sig (*p*):** Provides the *p*-value for the statistical test of significance (i.e., whether the variable is a significant predictor of compensation). *p*-values less than or equal to 0.05 are generally considered statistically significant.
  - **Valid n:** Provides the sample size of the analysis.
2. **Individual Factors:** Interpretation of the *Individual Factors* provides traditional *t*-Test statistics.
  - **b (*b*):** Provides an estimate of mean-group differences.
  - ***t*-value:** Provides the *t*-Test statistics.
  - **Sig (*p*):** Provides the *p*-value for the statistical test of significance (i.e., whether the model is a significant predictor of compensation). *p*-values less than or equal to 0.05 are generally considered statistically significant.

Because no explanatory variables are included in this model, the analyst should interpret the at-issue variable, which in this example is *Individual Factors/Minority*. The intercept [*Int(a)*] is largely ignored but may provide valuable insight for the more advanced analyst.

## Worksheet 9 – Stats(Exp Var):

The raw *regression* statistics are presented in the *Stats(Exp Var)* (“Explanatory Variables”) worksheet. This is the full regression model *with* all explanatory variables. Proper interpretation of the report requires a firm understanding of regression statistics. The report presents the statistics in two sections: *Overall Model* and *Individual Factors*. The columns in the worksheet are ordered, such that the at-issue *Test Group Variable* is presented first followed by all explanatory variables specified in the model. The components of this worksheet are presented below:

### Analysis Statistics: Regression with Explanatory Variables

Overall Model			Individual Factors					
	Statistics	Value	Statistics	Minority	TIJ	TIC	PPA	Int(a)
1	$R^2$	0.81	$b$	5493.63	-56.57	-3.26	5551.13	7667.90
	Sig( $p$ )	0.00	$t$ -value	9.07	-0.77	-0.05	36.71	19.82
	Valid n	363.00	Sig( $p$ )	0.00	0.44	0.96	0.00	0.00
	$R^2$	0.65	$b$	-10139.66	471.75	-4.57	1042.27	23564.23
	Sig( $p$ )	0.00	$t$ -value	-3.20	3.48	-0.04	4.38	15.81
	Valid n	27.00	Sig( $p$ )	0.00	0.00	0.97	0.00	0.00
3	$R^2$	0.91	$b$	-6234.18	122.30	1236.39	2359.70	30242.01
	Sig( $p$ )	0.00	$t$ -value	-1.26	0.94	5.26	8.32	22.29
	Valid n	87.00	Sig( $p$ )	0.21	0.35	0.00	0.00	0.00

1. Overall Model: Interpretation of the *Overall Model* statistics provides insight into the strength of the specified model.
  - R-Sqr ( $R^2$ ): Provides an estimate of effect size (i.e., the amount of variance accounted for by the model).
  - Sig ( $p$ ): Provides the  $p$ -value for the statistical test of significance (i.e., whether the model is a significant predictor of compensation).  $p$ -values less than or equal to 0.05 are generally considered statistically significant.
  - Valid n: Provides the sample size of the analysis.
2. Individual Factors: Interpretation of the *Individual Factors* provides traditional regression-coefficient statistics.
  - b ( $b$ ): Provides the unstandardized regression-coefficients; they can be used to determine the directionality and the strength of the relationship between the specific variable and compensation.
  - $t$ -value: Provides the  $t$ -Test statistics.
  - Sig ( $p$ ): Provides the  $p$ -value for the statistical test of significance (i.e., whether the model is a significant predictor of compensation).  $p$ -values less than or equal to 0.05 are generally considered statistically significant.

## **Worksheet 10 – Compensation Analysis:**

The *Comp Analysis* worksheet provides a powerful summary of the compensation analysis in a very compact and efficient format. At a glance, the analyst can identify which SSEGs were significant and the severity of the pay disparity. The components of this worksheet are presented below:

### Compensation Analysis Results

Job Group	Data	Focal	Reference	Grand Total	t-Impact/ Diff. (\$)	Reg-Impact/ Diff. (\$)	t-Test p-Value	Reg p-Value
1	Count of Minority	276	87	363	Focal	Focal	0.00	0.00
	Average of Salary	\$18,414.38	\$22,230.11	\$19,328.90	\$3,815.73	\$5,493.63		
	Average of PPA	2.0	1.8	1.9				
	Average of TIC	4.7	17.1	7.7				
	Average of TIJ	4.6	16.9	7.5				
2	Count of Minority	14	13	27	Focal	Reference	0.03	0.00
	Average of Salary	\$28,211.36	\$30,175.77	\$29,157.19	\$1,964.41	\$10,139.66		
	Average of PPA	3.2	3.0	3.1				
	Average of TIC	11.1	28.2	19.4				
	Average of TIJ	2.9	29.2	15.5				
3	Count of Minority	80	7	87	Focal	Reference	0.00	0.21
	Average of Salary	\$49,314.03	\$74,279.14	\$51,322.71	\$24,965.12	\$6,234.18		
	Average of PPA	4.2	5.7	4.4				
	Average of TIC	6.8	28.1	8.5				
	Average of TIJ	5.2	16.3	6.1				

The sample report above indicates:

- JobGroup 1: The *t*-Test indicates the pay disparity negatively impacts Focal members (\$3,815.73) and after controlling for the explanatory variables, the pay disparity is even greater (\$5,493.63).
- JobGroup 2: The *t*-Test indicates the pay disparity negatively impacts Focal members, but after controlling for the explanatory variables, the pay disparity is actually negatively impacting Reference members. This is not as uncommon as one would believe.
- JobGroup 3: The *t*-Test indicates the pay disparity significantly negatively impacts Focal members, but after controlling for explanatory variables, the pay disparity is no longer significant ( $p = 0.21$ ).

Each column of the report is discussed in detail below:

1. Columns left of the *Data* column are SSEG identifiers.
2. Data: Indicates the statistics that are reported in the *Focal*, *Reference*, and *Grand Total* columns.
3. Focal/Reference/Grand Total: These are the default headers. Analysts who customize the *Report Setup* can expect to see the headers that they specify.
4. *t*-Impact/Difference (\$): Provides results of the *t*-Test analyses, including the negatively impacted group and the raw mean difference between the two groups.
5. Reg-Impact/Difference (\$): Provides results drawn from the regression analyses, including the negatively impacted group and the *adjusted* mean difference between the two groups after controlling for the explanatory variables.
6. *t*-Test *p*-value: Provides the *p*-value for the *t*-Test (i.e., whether the model is a significant predictor of compensation absent any explanatory variables). *p*-values less than or equal to 0.05 are generally considered statistically significant.
7. Reg *p*-value: Provides the *p*-value for the multiple regression analysis (i.e., whether the protected variable is a significant predictor of compensation after controlling for any explanatory variables). *p*-values less than or equal to 0.05 are generally considered statistically significant.

If *NC* is indicated in the results, it means that the statistics could not be calculated either due to small sample size or because there were no members of one of the comparison groups. This is a useful report that may be printed and used during discussions with management.

## **Worksheet 11 – Liabilities:**

In the typical analysis, *Liabilities* is the last worksheet. This report provides an estimate of the additional pay that would need to be allocated to specific employees to reduce observed pay disparities to the desired standard deviation (SD) difference (default is 1-SD). Computed liabilities are presented in the last two columns of this report. The components of this worksheet are presented below:

---

***WARNING: Do not make adjustments to employee compensation without first determining whether other factors could be influencing pay in a legitimate manner and that the multiple regression model itself is valid and reliable.***

---

### Liabilities Report

Job Group	Emp ID	Salary	PPA	TIC	TIJ	Minority	t-Liable	R-Liable
1	272	\$34,787	6.0	8.0	7.0	0	\$0.00	\$5,856.54
	218	\$41,500	7.0	4.0	6.0	0	\$0.00	\$5,840.02
	72	\$30,415	5.0	3.0	4.0	0	\$0.00	\$5,753.71
	299	\$16,130	2.0	11.0	8.0	0	\$4,194.73	\$5,724.33
	279	\$16,113	2.0	7.0	8.0	0	\$4,225.94	\$5,702.30
	207	\$16,002	2.0	6.0	7.0	0	\$4,429.77	\$5,658.23
	204	\$16,048	2.0	9.0	7.0	0	\$4,345.30	\$5,638.03
	393	\$16,073	2.0	10.0	7.0	0	\$4,299.39	\$5,601.30
	441	\$16,389	2.0	9.0	8.0	0	\$3,719.14	\$5,531.53
	55	\$16,512	2.0	11.0	8.0	0	\$3,493.28	\$5,529.69
	355	\$16,573	2.0	9.0	8.0	0	\$3,381.26	\$5,515.00
	150	\$16,638	2.0	11.0	8.0	0	\$3,261.91	\$5,461.75
	201	\$16,113	2.0	5.0	6.0	0	\$4,225.94	\$5,452.57

1. t-Liable(\$): Provides computed liability, by employee, based on t-Test model.
2. R-Liable(\$): Provides computed liability, by employee, based on regression model.

If “Small N” is listed in the results, it indicates that the number of employees in the group was too small to make the necessary calculations. Values are only listed when an employee’s pay is sufficiently below their expected salary to warrant further consideration for a pay adjustment (based upon the analysts chosen level of allowable disparity). The other columns in this report are presented to assist the analyst in interpreting the liabilities should a cohort analysis be necessary. Note that the R-liabilities are sorted in descending order.

If a warning message appears on the header of this report stating “Warning: X cases were excluded from liability computations. Potential excluded liabilities is: \$X” a problem has occurred with the analysis where some data for one or more explanatory variables for the at-issue group is missing. When this error appears,

Full Information Maximum Likelihood (FIML) statistical modeling should be used to impute data for the missing records. Call BCG for assistance.

---

***Important Note: COMPARE is a powerful and comprehensive statistical tool that takes advantage of the advanced statistical technique known as multiple regression analysis to explore company pay practices. It is particularly useful in determining whether differences in pay are due to potential discrimination or due to legitimate, job-related factors. There is always a danger when automating such a powerful technique that steps will not be taken to ensure that the analyses comply with good scientific practice. Therefore, Biddle Consulting Group, Inc. strongly encourages users who are not highly experienced with advanced statistics to seek some level of oversight to ensure that the results of analyses are valid, reliable, and defensible.***

***Those who use COMPARE do so at their own risk and Biddle Consulting Group, Inc. makes no warranties, express or implied, with respect to the findings or liabilities resulting from its use.***

---

---

<sup>1</sup> When data has certain ordered properties, it may be coded into numeric format (e.g., Best=2, Acceptable=1, Worst=0). This is an advanced procedure and requires an analyst with sufficient experience.

<sup>2</sup> This essentially means that employees are grouped so that they are performing essentially the same duties, have the same minimum qualifications, require the same levels of supervision and have the same levels of responsibility.

<sup>3</sup> This is not a statistical test; it only provides a high-level indicator for potential disparity and should only be used as a guide.

<sup>4</sup> COMPARE tests for interactions at the aggregated-level—across all SSEGs. Although COMPARE is equipped to evaluate interactions for each SSEG, the sample-size requirements to test for interactions require a more conservative implementation. At the aggregated-level, COMPARE is most capable of testing and identifying potential interactions.

