Public Use Data File Documentation for the Study *Using Administrative Data to Address Federal Contractor Violations of Equal Employment Opportunity Laws*

July 5, 2013

Caroline Massad Francis
Nan Maxwell

MATHEMATICA
Policy Research

Public Use Data File Documentation for the Study *Using Administrative Data to Address Federal Contractor Violations of Equal Employment Opportunity Laws*

July 5, 2013

Caroline Massad Francis
Nan Maxwell

MATHEMATICA
Policy Research

CONTENTS

TABLES

INTRODUCTION

This documentation describes the data collected and used for the Administrative and Survey Data Research and Analysis (ADRA) study conducted by Mathematica Policy Research for the Office of Federal Contract Compliance Programs (OFCCP) at the U.S. Department of Labor (DOL). This study was funded by DOL's Chief Evaluation Office under contract #DOLF119432397. The purpose of the study was to provide a better understanding of data available to examine violations and repeat violations of equal employment opportunity (EEO) laws by federal contractors. Mathematica used administrative data from OFCCP on contractors reviewed between fiscal years 2003 and 2012 and estimated (1) violation rates among the federal contractors OFCCP investigates and the characteristics of violators, (2) re-violation rates among the federal contractors OFCCP investigates and the characteristics of re-violators, and (3) factors associated with a contractor violating or re-violating EEO laws.

Results from the study are summarized in Maxwell et al. (2013)[1] and were based on information from the following sources:[2]

1. **The OFCCP Information System (OFIS)** records the details of reviews of federal contractors and includes information about the contractor being reviewed and violations found during the review.

2. **Conciliation agreement and consent decree (CA/CD)** documents describe the terms of the remedies for violations found in OFCCP reviews.

3. **Bureau of Labor Statistics (BLS)** data for each U.S. county during each year capture the economic conditions in which contractors operated at the time of the review. We link local unemployment rates from this data source to illustrate that linking OFIS to other databases could enhance its value in research.

The public use data file for the study is **adra_ofccp_pudf.** It contains 46,198 records, with a single record for each review (a review is equivalent to a case) undertaken between fiscal years 2003 and 2012. The data file contains several interrelated data entities: contractors, reviews, violations, and remedies. A contractor in the data could be associated with one or more reviews; each review could find zero or more violations; and each violation could be associated with zero or more remedies. Maxwell et al. (2013) describe these relationships in detail and how the relationships are structured on the data file.

In Section I of this documentation, we present selected information about each input data source, a fuller discussion of which is available in Maxwell et al. (2013). In Section II, we itemize the contents of the data file, and, in Section III, we present summary statistics for the data file.

---

[1] Maxwell, Nan, Aravind Moothy, Caroline Massad Francis, and Dylan Ellis. "Using Administrative Data to Address Federal Contractor Violations of Equal Employment Opportunity Laws." Report submitted to the U.S. Department of Labor, Office of the Assistant Secretary for Administration and Management. Oakland, CA: Mathematica Policy Research, 2013.

[2] Results from the study were also based on press releases issued by DOL about major violations. However, we exclude press release information from the public use file in order to protect the identity of contractators.

I. INFORMATION ABOUT INPUT DATA SOURCES

The 1960s were a fertile time for improving the legal rights of Americans to equal employment opportunity (EEO), as the Equal Pay Act, Civil Rights Act, and specific antidiscrimination requirements for federal contractors were established. Two agencies were charged with enforcing EEO requirements: (1) the U.S. Equal Employment Opportunity Commission (EEOC), which has a general mission covering all employers, workers, and job applicants, and (2) OFCCP, an agency within the U.S. Department of Labor (DOL) that enforces EEO and affirmative action requirements among firms that contract with the federal government.

To carry out its mission, OFCCP regularly reviews contractors' personnel data and policies and responds to complaints. If OFCCP finds evidence of a violation of EEO legislation or affirmative action requirements during a review, it attempts to work with the contractor to define actions the contractor will take to stop the violation, prevent its recurrence, and, if applicable, compensate victims. OFCCP may also pursue termination of a contract or debarment from future federal contracts if a contractor fails to agree to take acceptable remedial actions.

The **adra_ofccp_pudf** file was developed to use in research that investigates whether the tools and remedies that OFCCP uses to ensure that contractors comply with EEO and affirmative action requirements help to correct past behavior and deter repeat violations. It draws on information from both OFCCP and BLS.

A. Data from OFCCP

OFIS records and CA/CD documents were provided to Mathematica by OFCCP for the period between October 26, 2012 and December 21, 2012.[3] OFCCP provided (1) information from OFIS in three separate files and (2) the CA/CDs in folders, each labeled with a fiscal year. Mathematica merged these data with BLS county unemployment rates. Table I.1 provides details of each data source, including content, format, and number of records.

Mathematica cleaned and processed information from these files in four ways, listed here. These procedures are detailed in Appendices A and B in Maxwell et al. (2013), which researchers should read before using the data.

1. We created a random identification number for each review, establishment, and parent firm and removed or modified data elements that could be used to determine the review and contractor identity. This process removed geographic identifiers, including state, city, and zip code, from the file and replaced the date of review closure with the fiscal year in which the review closed.

2. We extracted remedy data from CA/CD documents to incorporate into the OFIS file using an optical character recognition software program to convert each document file into a text file that could be searched and a list of keywords associated with each remedy type to identify the types of remedies in each document.

---

[3] All information about Mathematica was redacted.

Table I.1. Description of the Files from OFCCP

Source	Component	Contents	Format	Number of Records
OFCCP				
OFIS	Compliance Evaluation	Information for each review closed between fiscal years 2003 and 2012 that did not originate from a complaint	Excel file. Each record corresponds to a single review.	40,502 reviews
OFIS	Complaint Investigations	Reviews and violation information for reviews closed during fiscal years 2003 to 2012 that originated from a complaint	Excel file. Each record corresponds to a single review.	5,696 reviews
OFIS	Compliance Evaluation Violations	Information about the violations found in the Compliance Evaluation Reviews file	Excel file. Each record corresponds to a single violation (a single review could have more than one violation).	17,293 violations
CA/ CD	Conciliation Agreement/ Consent Decrees	Documents containing remedies for violations found in Compliance Evaluation Reviews and Complaint Investigations	15 compressed folders, each labeled with a fiscal year. Folders contained 2,650 documents, most corresponding to a single review with at least one violation.	2,650 documents
BLS				
LAUS	Unemployment Rates	Average annual unemployment rates in each U.S. county	Text file. Each record corresponds to a county in a particular year.	3,218 records per year (average)

BLS = Bureau of Labor Statistics; CA/CD = Conciliation Agreement/Consent Decree; DOL = U.S. Department of Labor; LAUS = Local area unemployment statistics; OFIS = OFCCP Information System.

3. We developed an algorithm to more accurately determine whether a review resulted in violations and, if it did, whether the violations were discriminatory or only technical.

4. We recoded some information from OFIS if it was not valid in the original files (for example, binary indicator variables for each violation code were coded as missing if a review had a violation but no violation codes were coded as "1" in OFIS, and they were coded as legitimate skips if OFCCP found no violation in the review). Table III.1 at the end of this document contains information about missing data and legitimate skips for each variable.

In addition, we capped the number of employees, number of victims, financial agreement amount, and unemployment rate variables in the public use file, protect contractors' identities.

B. Data from BLS

We mapped claimants' zip codes to counties in 2010[4] in order to measure local area unemployment rates.[5] If the zip code boundary was the same as or within the boundary of a county, we assigned the zip code the unemployment rate of that county. If the zip code boundary mapped to more than one county, we assigned the zip code the weighted average of the county unemployment rates. The weight used in this average was the county population in 2010.

---

[4] We used the MABLE/Georcorr2K conversion engine [http://mcdc2.missouri.edu/websas/geocorr2k.html] to identify how the zip codes overlapped.

[5] See Bureau of Labor Statistics Local Area Unemployment Statistics [http://www.bls.gov/lau/data.htm].

II. CONTENT OF THE FINAL DATA FILE

The data file **adra_ofccp_pudf** contains review-level information for reviews of contractors that OFCCP performed between fiscal years 2003 and 2012. Table II.1 lists the variable names, variable labels, data types (categorical, continuous, or string), and value ranges (for nonstring variables) in the file. The variables are grouped according to the key constructs used in Maxwell et al. (2013):

- Identifier variables

- Contractor characteristics

- Investigative tools

- Violation characteristics

- Remedies

- External environment

- Original and source variables

Within each group, variables are ordered so that they follow the order of variables in the tables in Maxwell et al. (2013). The category "original and source variables" comprises variables not used in that study, but that other researchers may find useful. They are (1) violation type and violation basis, as they appeared in OFIS before we applied an algorithm for correcting violation types; (2) remedies as they appeared when first extracted from the CA/CD file, before recoding based on information about remedies in OFIS; and (3) a flag that indicates which reviews came from the compliance evaluation file and which from the complaint investigations file.

In describing value ranges in Table II.1, we generally list the minimum and maximum values for continuous and categorical variables.[6] Definitions of the values of the categorical variable are not listed here, but are instead presented as part of the tabulations in Section III. In a final note, as in our analysis in Maxwell et al. (2013), we group all but the most common industry and violation codes into an "other" category to protect contractor identification.

---

[6] The exception to this rule is naics_cd_2; we present a list of discrete values because we use the original 2-digit North American Industry Classification System (NAICS) codes. For other numeric categorical variables, values comprise all whole numbers within the range given.

Table II.1. Content of the Final Data File

Variable Name	Variable Label	Type	Value Range
Identifier Variables			
mpr_review_id	Review ID number	String	n.a.
mpr_e_id	Establishment ID number	String	n.a.
mpr_p_id	Parent ID number	String	n.a.
Contractor Characteristics			
Size			
size	Number of employees	Numeric (continuous)	[1, 1,000]
p_size	Parent number of employees (imputed)	Numeric (continuous)	[4, 1,000]
Industry			
naics_cd_2	Industry (2-digit NAICS code)	Numeric (categorical)	[31, 54, 56, 62, 100][a]
Parent			
has_parent	Review includes P-EEO1 or parent name	Numeric (categorical)	[0,1]
Investigative Tool			
inv_tool	Investigative tool	Numeric (categorical)	[1,2]
Violation Characteristics			
Violation Type			
viol_type	Violation type	Numeric (categorical)	[0,2]
Violation Code			
viol_cd_8	Hiring violation	Numeric (categorical)	[0,1]
viol_cd_16	Selection/testing violation	Numeric (categorical)	[0,1]
viol_cd_17	Salary violation	Numeric (categorical)	[0,1]
viol_cd_22	Systemic discrimination violation	Numeric (categorical)	[0,1]
viol_cd_100	Other discrimination violation	Numeric (categorical)	[0,1]
Violation Basis			
viol_b_sex	Violation basis sex	Numeric (categorical)	[0,1]
viol_b_col	Violation basis color	Numeric (categorical)	[0,1]
viol_b_vet	Violation basis veteran	Numeric (categorical)	[0,1]
viol_b_dsb	Violation basis disability status	Numeric (categorical)	[0,1]
viol_b_nor	Violation basis national origin	Numeric (categorical)	[0,1]
viol_b_rlg	Violation basis religion	Numeric (categorical)	[0,1]
Victims			
victims	Number of victims	Numeric (continuous)	[1, 200]
Remedies/Sanctions			
has_cacd	Has a CA/CD	Numeric (categorical)	[0,1]
remedy_emp	Employment remedy	Numeric (categorical)	[0,1]
remedy_fin	Financial remedy	Numeric (categorical)	[0,1]
financial_amt	Financial agreement amount in dollars	Numeric (continuous)	[1, 500,000]
remedy_org	Organizational change remedy	Numeric (categorical)	[0,1]
has_pr	Has a press release	Numeric (categorical)	[0,1]
External Environment			
Region			
region	Regional office	Numeric (categorical)	[1, 7]
p_region	Parent region (imputed)	Numeric (continuous)	[1, 6]
Year Review Closed			
year	Fiscal year review closed	Numeric (categorical)	[2003, 2012]
v_year	Year review with violation closed	Numeric (categorical)	[2003, 2012]
Local Area Unemployment Rate			
unemp	Local area unemployment rate	Numeric (continuous)	[1.9, 10.0]
p_unemp	Parent local area unemployment rate	Numeric (continuous)	[1.9, 10.0]

Variable Name	Variable Label	Type	Value Range
Original and Source Variables			
Variables from OFIS			
Violation Type			
ofis_discrim	Violation type (OFIS)	Numeric (categorical)	[0, 2]
Violation Basis			
ofis_viol_b_sex	Violation basis sex (OFIS)	Numeric (categorical)	[0,1]
ofis_viol_b_col	Violation basis color (OFIS)	Numeric (categorical)	[0,1]
ofis_viol_b_vet	Violation basis veteran (OFIS)	Numeric (categorical)	[0,1]
ofis_viol_b_dsb	Violation basis disability status (OFIS)	Numeric (categorical)	[0,1]
ofis_viol_b_nor	Violation basis national origin (OFIS)	Numeric (categorical)	[0,1]
ofis_viol_b_rlg	Violation basis religion (OFIS)	Numeric (categorical)	[0,1]
Variables from CA/CDs			
ca_remedy_emp	Employment remedy (CA/CD)	Numeric (categorical)	[0,1]
ca_remedy_fin	Financial remedy (CA/CD)	Numeric (categorical)	[0,1]
ca_remedy_org	Organizational change remedy (CA/CD)	Numeric (categorical)	[0,1]
ca_remedy_missing	Missing remedies in CA/CD	Numeric (categorical)	[0,1]
Flags			
complaints_flag	Review is from complaint investigations file	Numeric (categorical)	[0,1]

Notes: Value ranges are presented only for numeric categorical and continuous variables.

[a] 100 is the code assigned to industries other than the four most common.

n.a. = not applicable.

III. SUMMARY STATISTICS

This section presents summary statistics for the data file. The following information is presented for all variables:

- The variable's name, label, and type (categorical or continuous), which replicates the information in Table II.1

- The number of distinct values

- The number of observations with missing values relative to the total number of observations, and the number of observations with legitimate skips relative to the total number of observations[7]

The summary statistics also include frequency and percentage tabulations for categorical variables with up to 10 values, which is the number of categories for the year variable. These tabulations include both the values and the value labels to facilitate interpretation of the resulting numbers. For continuous variables, the summary statistics instead include the mean, standard deviation, and values at the 10th, 25th, 50th, 75th, and 90th percentiles.

---

[7] Original OFIS variables have no legitimate skips in this data file because OFIS did not include codes for legitimate skips.

Table III.1. Summary Statistics for adra_ofccp_pudf

Variable name: mpr_review_id

Variable label:	Review ID Number
Variable type:	String

Number of distinct values:	46,198
Total number of cases:	46,198
Cases with missing values / total number of cases (percent):	0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):	0 / 46,198 (0.0%)

Tabulation suppressed due to large number of distinct values.

Variable name: mpr_e_id

Variable label:	Establishment ID Number
Variable type:	String

Number of distinct values:	31,498
Total number of cases:	46,198
Cases with missing values / total number of cases (percent):	8,412 / 46,198 (18.2%)
Cases with legitimate skips / total number of cases (percent):	0 / 46,198 (0.0%)

Tabulation suppressed due to large number of distinct values.

Variable name: mpr_p_id

Variable label:	Parent ID Number
Variable type:	String

Number of distinct values:	7,938
Total number of cases:	46,198
Cases with missing values / total number of cases (percent):	27,381 / 46,198 (59.3%)
Cases with legitimate skips / total number of cases (percent):	0 / 46,198 (0.0%)

Tabulation suppressed due to large number of distinct values.

Variable name: size

Variable label:	Number of employees
Variable type:	Numeric (continuous)

Number of distinct values:	1,000
Total number of cases:	46,198
Cases with missing values / total number of cases (percent):	8,230 / 46,198 (17.8%)
Cases with legitimate skips / total number of cases (percent):	0 / 46,198 (0.0%)

Mean	Standard Deviation	Minimum	Maximum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
333.6	280.0	1.0	1,000.0	72.0	142.0	232.0	434.0	872.0

Variable name: p_size

Variable label: Parent number of employees (imputed)
Variable type: Numeric (continuous)

Number of distinct values: 936
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   12,725 / 46,198 (27.5%)
Cases with legitimate skips / total number of cases (percent):   15,018 / 46,198 (32.5%)

Mean	Standard Deviation	Minimum	Maximum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
733.9	347.8	4.0	1,000.0	172.0	376.0	1,000.0	1,000.0	1,000.0

Variable name: naics_cd_2

Variable label: Industry (2-digit NAICS code)
Variable type: Numeric (categorical)

Number of distinct values: 5
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   6,208 / 46,198 (13.4%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
100	Other industry	16,335	35.4%
31	Manufacturing	13,550	29.3%
54	Professional/scientific/technical	4,534	9.8%
56	Administrative support	2,360	5.1%
62	Health care and social assistance	3,211	7.0%

Variable name: has_parent

Variable label: Review includes P-EEO1 or parent name
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	15,018	32.5%
1	Yes	31,180	67.5%

Variable name: inv_tool

Variable label: Investigative tool
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   2,646 / 46,198 (5.7%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
1	Selection system review	31,422	68.0%
2	Non-selection system review	12,130	26.3%

Variable name: viol_type

Variable label: Violation type
Variable type: Numeric (categorical)

Number of distinct values: 3
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	None	37,733	81.7%
1	Discrimination	864	1.9%
2	Technical only	7,601	16.5%

Variable name: viol_cd_8

Variable label: Hiring violation
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   74 / 46,198 (0.2%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	7,809	16.9%
1	Yes	582	1.3%

Variable name: viol_cd_16

Variable label: Selection/testing violation
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   74 / 46,198 (0.2%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	8,345	18.1%
1	Yes	46	0.1%

Variable name: viol_cd_17

Variable label: Salary violation
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   74 / 46,198 (0.2%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	8,247	17.9%
1	Yes	144	0.3%

Variable name: viol_cd_22

Variable label: Systemic discrimination violation
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   74 / 46,198 (0.2%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	7,929	17.2%
1	Yes	462	1.0%

Variable name: viol_cd_100

Variable label: Other discrimination violation
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   74 / 46,198 (0.2%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	8,203	17.8%
1	Yes	188	0.4%

Variable name: viol_b_sex

Variable label: Violation basis sex
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   68 / 46,198 (0.1%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	2,040	4.4%
1	Yes	6,357	13.8%

Variable name: viol_b_col

Variable label: Violation basis color
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   68 / 46,198 (0.1%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	2,505	5.4%
1	Yes	5,892	12.8%

Variable name: viol_b_vet

Variable label: Violation basis veteran
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   68 / 46,198 (0.1%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	4,260	9.2%
1	Yes	4,137	9.0%

Variable name: viol_b_dsb

Variable label: Violation basis disability status
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   68 / 46,198 (0.1%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	5,342	11.6%
1	Yes	3,055	6.6%

Variable name: viol_b_nor

Variable label: Violation basis national origin
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   68 / 46,198 (0.1%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	7,442	16.1%
1	Yes	955	2.1%

Variable name: viol_b_rlg

Variable label: Violation basis religion
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   68 / 46,198 (0.1%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	7,892	17.1%
1	Yes	505	1.1%

Variable name: victims

Variable label: Number of victims
Variable type: Numeric (continuous)

Number of distinct values: 168
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   1 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   45,334 / 46,198 (98.1%)

Mean	Standard Deviation	Minimum	Maximum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
93.6	83.6	1.0	200.0	2.0	6.0	71.0	200.0	200.0

Variable name: has_cacd

Variable label: Has a CA/CD
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	6,795	14.7%
1	Yes	1,670	3.6%

Variable name: remedy_emp

Variable label: Employment remedy
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   7,737 / 46,198 (16.7%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	384	0.8%
1	Yes	344	0.7%

Variable name: remedy_fin

Variable label: Financial remedy
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   868 / 46,198 (1.9%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	6,627	14.3%
1	Yes	970	2.1%

Variable name: financial_amt

Variable label: Financial agreement amount in dollars
Variable type: Numeric (continuous)

Number of distinct values: 571
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   15 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Mean	Standard Deviation	Minimum	Maximum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
26,657.1	100,470.0	0.0	500,000.0	0.0	0.0	0.0	0.0	6,200.0

Variable name: remedy_org

Variable label: Organizational change remedy
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   7,737 / 46,198 (16.7%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
0	No	182	0.4%
1	Yes	546	1.2%

Variable name: region

Variable label: Regional office
Variable type: Numeric (categorical)

Number of distinct values: 7
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
1	Northeast	5,899	12.8%
2	Mid-Atlantic	6,459	14.0%
3	Southeast	10,041	21.7%
4	Midwest	8,335	18.0%
5	Southwest	7,156	15.5%
6	Pacific	7,921	17.1%
7	National Office	387	0.8%

Variable name: p_region

Variable label: Parent region (imputed)
Variable type: Numeric (categorical)

Number of distinct values: 6
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   9,995 / 46,198 (21.6%)
Cases with legitimate skips / total number of cases (percent):   15,018 / 46,198 (32.5%)

Value	Label	Frequency	Percentage
1	Northeast	4,131	8.9%
2	Mid-Atlantic	2,840	6.1%
3	Southeast	3,800	8.2%
4	Midwest	4,873	10.5%
5	Southwest	2,829	6.1%
6	Pacific	2,712	5.9%

Variable name: year

Variable label: Fiscal year review closed
Variable type: Numeric (categorical)

Number of distinct values: 10
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
2003	2003	1,094	2.4%
2004	2004	7,316	15.8%
2005	2005	3,434	7.4%
2006	2006	4,533	9.8%
2007	2007	5,439	11.8%
2008	2008	4,939	10.7%
2009	2009	4,429	9.6%
2010	2010	5,550	12.0%
2011	2011	4,676	10.1%
2012	2012	4,788	10.4%

Variable name: v_year

Variable label: Year review with violation closed
Variable type: Numeric (categorical)

Number of distinct values: 10
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   37,733 / 46,198 (81.7%)

Value	Label	Frequency	Percentage
2003	2003	1,079	2.3%
2004	2004	645	1.4%
2005	2005	475	1.0%
2006	2006	565	1.2%
2007	2007	653	1.4%
2008	2008	698	1.5%
2009	2009	744	1.6%
2010	2010	1,040	2.3%
2011	2011	1,178	2.5%
2012	2012	1,388	3.0%

Variable name: unemp

Variable label: Local area unemployment rate
Variable type: Numeric (continuous)

Number of distinct values: 4,536
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   7,738 / 46,198 (16.7%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Mean	Standard Deviation	Minimum	Maximum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
6.3	2.2	1.9	10.0	3.7	4.5	5.8	8.1	10.0

Variable name: p_unemp

Variable label: Parent local area unemployment rate
Variable type: Numeric (continuous)

Number of distinct values: 1,221
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   14,180 / 46,198 (30.7%)
Cases with legitimate skips / total number of cases (percent):   15,018 / 46,198 (32.5%)

Mean	Standard Deviation	Minimum	Maximum	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
6.3	2.1	1.9	10.0	3.9	4.7	5.8	7.7	9.8

Variable name: ofis_discrim

Variable label: Violation type (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 3
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	N	39,438	85.4%
1	S	863	1.9%
2	T	5,897	12.8%

Variable name: ofis_viol_b_sex

Variable label: Violation basis sex (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	39,676	85.9%
1	Yes	6,522	14.1%

Variable name: ofis_viol_b_col

Variable label: Violation basis color (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	40,149	86.9%
1	Yes	6,049	13.1%

Variable name: ofis_viol_b_vet

Variable label: Violation basis veteran (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	41,859	90.6%
1	Yes	4,339	9.4%

Variable name: ofis_viol_b_dsb

Variable label: Violation basis disability status (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	43,033	93.1%
1	Yes	3,165	6.9%

Variable name: ofis_viol_b_nor

Variable label: Violation basis national origin (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	45,206	97.9%
1	Yes	992	2.1%

Variable name: ofis_viol_b_rlg

Variable label: Violation basis religion (OFIS)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	No	45,664	98.8%
1	Yes	534	1.2%

Variable name: ca_remedy_emp

Variable label: Employment remedy (CA/CD)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   942 / 46,198 (2.0%)
Cases with legitimate skips / total number of cases (percent):   44,528 / 46,198 (96.4%)

Value	Label	Frequency	Percentage
0	No	384	0.8%
1	Yes	344	0.7%

Variable name: ca_remedy_fin

Variable label: Financial remedy (CA/CD)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   942 / 46,198 (2.0%)
Cases with legitimate skips / total number of cases (percent):   44,528 / 46,198 (96.4%)

Value	Label	Frequency	Percentage
0	No	359	0.8%
1	Yes	369	0.8%

Variable name: ca_remedy_org

Variable label: Organizational change remedy (CA/CD)
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   942 / 46,198 (2.0%)
Cases with legitimate skips / total number of cases (percent):   44,528 / 46,198 (96.4%)

Value	Label	Frequency	Percentage
0	No	182	0.4%
1	Yes	546	1.2%

Variable name: ca_remedy_missing

Variable label: Missing remedies in CA/CD
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   44,528 / 46,198 (96.4%)

Value	Label	Frequency	Percentage
0	Not Missing	728	1.6%
1	Missing	942	2.0%

Variable name: complaints_flag

Variable label: Review is from complaint investigations file
Variable type: Numeric (categorical)

Number of distinct values: 2
Total number of cases: 46,198
Cases with missing values / total number of cases (percent):   0 / 46,198 (0.0%)
Cases with legitimate skips / total number of cases (percent):   0 / 46,198 (0.0%)

Value	Label	Frequency	Percentage
0	Compliance evaluation file	40,502	87.7%
1	Complaint investigations file	5,696	12.3%

MATHEMATICA
Policy Research