

Description of the Data:

The data covers the years 2003 to 2013.

## Variables:

1. year: payroll year
2. id: unique individual identifier (this is just a masked id based on employee's SIN/SSN)
3. gender
4. job title
5. location: employee's office/plant is Canada or USA
6. location\_other: Example is situations that can arise from raw records—some employees had a different flag for their location. You will have to figure out how to code all of this.
7. salary
8. grade: Pay grade, there are 7 grades, 2-8 (1 is an open grade). See below for details
9. rating: Performance rating (1: requires improvement, 2: below expectations, 3: meets expectations, 4: exceeds expectations)
10. hiredate: Date of hire
11. minority flag: Self-reported member of visible minority (already coded as =1 for minority group, =0 for not)

Questions:

1. Equity Audit: Salary

$$\text{Salary} = \alpha + \beta_1 \text{Gender} + \beta_2 \text{Minority} + \gamma \mathbf{X} + \varepsilon$$

Question 1 requires estimating the equation above where gender is a dummy for gender and analogously for members of a visible minority. The key thing to do on question 1 is to decide which employee characteristics you will include in vector  $\mathbf{X}$ . Of course, you are welcome to estimate more than 1 model (that's not required). Once you have decided on your salary model(s), estimate it and interpret all coefficients with of course a particular focus on addressing the question of whether the company has human rights disparate impact risks. Recall from class (Nov 12<sup>th</sup> ), that standard (economics) statistical significance (10%, 5%, 1%) thresholds apply for litigation; however, you also want to think of this as a general risk management exercise as well (I'll leave that for you to think about!).

**Grading for all three of these outcomes (see below for the others) is based on both the model you estimate (i.e., does it make sense; is there an obvious characteristic you left out—has to be obvious) and then your brief write-up of the results. Write-up can be informal, bullet-point—like a memo-to-file. 1 page for each outcome is likely sufficient (but don't worry if you do more than that)**

2. Equity Audit: Salary at Hire

$$SalaryAtHire = \alpha + \beta_1 Gender + \beta_2 Minority + \gamma X + \varepsilon$$

Same as above but for question 2 you are analyzing *only* the salary upon hire. Thus, for Question 2, we are working with a *sub-sample*: only employees in their first year at the company. Note that we cannot verify salary at hire for employees hired in 2001 and 2002 and thus those observations need to be dropped from the sample.

3. Promotions

$$Promoted = \alpha + \beta_1 Gender + \beta_2 Minority + \gamma X + \varepsilon$$

where Promoted =1 if the employee was promoted, =0 for if the individual is not promoted. For this analysis, we will be creating a cross-sectional database at the individual level; alternatively stated, our regression will be estimated with collapsed data (collapsed to the individual level as opposed to a pooled cross-sectional-time series data set). If you are doing this in Excel, you will have to create a new dataset by hand (at least as far I am aware—if you figure out a way to do using macros or something let me know!). Thus, promoted will =1 if a given individual was ever promoted (and thus=0 if never promoted).

Extra step: For the promotions question, there is an extra question to address in your memo—what drawbacks do you see with the promotions model? When answering this question, keep in mind the goal here really is the  $\beta_1$  and  $\beta_2$  coefficients. That said, I will also accept any interesting comments about promotions modelling (so feel free to think outside of the box using your knowledge from Project 1 about promotions generally).

**Pay System Details:**

<b>2003 to 2005</b>			
Grade	Min	Mid	Max
1			
2	30000	40000	50000
3	34500	46000	57500
4	39675	52900	66125
5	51578	68770	85963
6	61893	82524	103155
7	71177	94903	118629
8	81854	109138	136423

<b>2006 to 2007</b>			
Grade	Min	Mid	Max
1			
2	30450	40600	50750
3	35018	46690	58363
4	40270	53694	67117
5	52351	69802	87252
6	62821	83762	104702
7	72245	96327	120408
8	83081	110775	138469

<b>2008 to 2011</b>			
Grade	Min	Mid	Max
1			
2	31668	42224	52780
3	36418	48558	60697
4	41881	55841	69802
5	54445	72594	90742
6	65334	87112	108890
7	75135	100180	125225
8	86405	115206	144008

<b>2012 to 2013</b>			
Grade	Min	Mid	Max
1			
2	33726	44969	56211
3	38785	51714	64642
4	44603	59471	74339
5	57984	77312	96640
6	69581	92775	115968
7	80018	106691	133364
8	92021	122694	153368