Michael Frajman - mhf8621
CORE-GP 1011 – Statistical Methods
December 12, 2022

Memo: Statistical Findings for use in Safecorp Discrimination Lawsuit

Equal opportunity employment comes with equal opportunity to earn the same pay. Salary equity remains a goal to strive toward with the United States Bureau of Labor Statistics reporting in November 2021 that Black and Hispanic employees continue to be paid less than their White and Asian counterparts.[1] Several minority employees of Safecorp bank have filed a lawsuit alleging discrepancies in their pay compared to their non-minority co-workers. The following memo was prepared at the request of those complainants and their legal counsel and provides a statistical analysis of Safecorp's employee and salary data. The complainants seek to determine if they as minorities are being paid less than non-minorities at the company as well as find what other factors may be used to predict an employee's pay.

The Safecorp data includes the following information for each employee: (1) Annual salary, (2) Number of years spent working at the company, (3) Position at the company, (4) Status as a minority, (5) Sex, and (6) Highest grade level of education. The data was processed using the STATA statistical package. Table 1 included in appendix I of this memo summarizing how this information breaks down for minority and non-minority employees. Several tests for significance were performed to determine which factors were predictors for salary and if these factors are associated with minority status.

[1] US Bureau of Labor Statistics. (2021, November). *Labor force characteristics by race and ethnicity.* Retrieved December 10 2022 from https://www.bls.gov/opub/reports/race-and-ethnicity/2020/home.htm#:~:text=The%20labor%20force%20participation%20rate%20for%20Asian%20adult%20men%20(74.0,and%20Whites%20(56.8%20percent).

The first major question that was explored was **if salaries for minority employees were truly lower than those for non-minority employees**. Of Safecorp's 474 employees, 21.9% identify as minorities. **It was found that the average annual salary for a minority employee was $57427.88 while non-minorities made an average of $72046.62** (see table 1). It can be stated with a **95% degree of certainty, that the average salaries for non-minority employees were significantly greater than those of minority employees**.

After determining that average salary did in fact vary significantly by status as a minority, **statistical analysis was performed to determine which employee characteristics could best predict salary.** All five employee characteristics (number of years, position, minority status, sex and grade level obtained) were modelled to determine how much they affect annual salary of an employee. Individually, all five were found to be significant in predicting salary. Table 2 lists by how much each characteristic affected annual salary. Using the above mentioned five characteristics this model was able to predict 73% of the variation in employee salary to a significant degree of certainty. The model found that if an employee is a minority, they will make on average $4388.26 less than a non-minority employee each year. However, sex may be a stronger factor in predicting salary with, a woman predicted to make $10965.61 less on average than a man. Conversely, occupying a managerial position and having more education indicates higher salaries on average.

These five characteristics were then compared between minority and non-minority employees to see if there was significant variation. Referring in order to the categories in table 1 which compares characteristics of employees by minority status: (1) the average number of years minority and non-minority employees have spent working at Safecorp are similar with no

significant difference found; (2) significant association was found between a person's minority status and the job position they hold within the company with only 3.9% of minorities occupying managerial positions compared to a much larger 21.6% of non-minorities; (3) minority and non-minority employees have comparable breakdowns by sex as no significant association; and (4) a significant difference was found between the average grade level obtained of minority and majority employees even though the numbers seem close at first glance with minorities completing 12.77 grades of school on average compared to 13.69 for non minorities. The differences in education level between the groups and the smaller number of minorities in higher level job positions are therefore worth noting given how these factors can predict higher salary. It should be stated that **all findings do not imply that these factors necessarily cause or explain the discrepancies in average salary observed between minorities and non-minorities, correlation does not imply causation.**

To conclude, non-minority employees of Safecorp do earn a greater amount of money on average than minority employees that is statistically significant. All employee characteristics Safecorp tracks are significant predictors of salary. Of those characteristics, job position was found to be significantly associated with minority status and highest average grade level of education was found to significantly differ depending on minority status. Even with this knowledge, this analysis makes no determination that any of these factor's are causes for the differences in salary between employees. This memo only states that differences exist and can be used to make predictions about salary. Further analysis would need to be performed to state with more confidence what other factors may influence things like the amount of education or career advancement minorities receive as well as examining Safecorp's human resource practices.

Appendix I – Tables

Table 1 - Characteristics and means for minority and non-minority employees of Safecorp

Characteristic	Minority (n = 104)	Non-minority (n = 370)
Annual salary (mean)	$57427.88	$72046.62*
Years at company (mean)	15.95	17.15
Position (%)		
Clerical	83.7	74.6**
Custodial	12.5	3.8
Manager	3.9	21.6
Sex (%)		
Male	61.5	52.4
Female	38.5	47.6
Highest grade completed (mean)	12.77	13.69***

* p < .1 for an independent one-tailed samples test of the difference of means between minorities and non-minorities.
** p < .05 for a chi-square test of the association of this characteristic with minorities versus non-minorities comparison.
*** p < .05 for an independent two-tailed samples test of the difference of means between minorities and non-minorities.

Table 2 – The average variation in salary each characteristic predicts in the regression analysis of annual salary

Characteristic	Average effect on salary ($)
1 additional year at company	-212.20
Identifying as a "minority"	-4388.26
Identifying as "female"	-10965.61
1 additional grade of education	+2916.88
Having a managerial role	+44957.17
Having a clerical role	-8759.84

All characteristics were deemed significant with p <0.05

Appendix II – STATA Output

```
.
. *-------------------------------------------------------------
. *------------------EXPLORATORY ANALYSIS-----------------------
. * descriptive statistics used in table
. *-------------------------------------------------------------
.
. * frequency table for MINORITY, check if there are "missings"
. tabulate minority, missing

    Employee's |
      minority |
        status |      Freq.     Percent        Cum.
---------------+-----------------------------------
"Non-minority" |        370       78.06       78.06
    "Minority" |        104       21.94      100.00
---------------+-----------------------------------
         Total |        474      100.00

.
. * frequency table for SEX, check if there are "missings"
. tabulate sex, missing

 Employee's |
        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
     "Male" |        258       54.43       54.43
     Female |        216       45.57      100.00
------------+-----------------------------------
      Total |        474      100.00

.
. * two way table comparing SEX and MINORIY
. tabulate sex minority, column

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |  Employee's minority
Employee's |        status
       sex | "Non-mino  "Minority |     Total
-----------+----------------------+----------
    "Male" |       194         64 |       258
           |     52.43      61.54 |     54.43
-----------+----------------------+----------
    Female |       176         40 |       216
           |     47.57      38.46 |     45.57
-----------+----------------------+----------
     Total |       370        104 |       474
           |    100.00     100.00 |    100.00

.
. * two way table comparing POSITION and MINORITY
. tabulate position minority, column

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+
```

```
Employee's |
  position |  Employee's minority
        at |        status
  Safecorp | "Non-mino  "Minority |     Total
-----------+----------------------+----------
  Clerical |       276         87 |       363
           |     74.59      83.65 |     76.58
-----------+----------------------+----------
 Custodial |        14         13 |        27
           |      3.78      12.50 |      5.70
-----------+----------------------+----------
   Manager |        80          4 |        84
           |     21.62       3.85 |     17.72
-----------+----------------------+----------
     Total |       370        104 |       474
           |    100.00     100.00 |    100.00
```

```
.
. * comparing MINORITY by statistics of ANNUAL_SAL
. tabstat annual_sal, statistics( mean sd median count min max ) by(minority)

Summary for variables: annual_sal
Group variable: minority (Employee's minority status)

      minority |      Mean        SD       p50         N       Min       Max
---------------+--------------------------------------------------------------
"Non-minority" |  72046.62  36088.19     59850       370     31500    270000
    "Minority" |  57427.88  22843.28     53250       104     32700    200000
---------------+--------------------------------------------------------------
         Total |  68839.14  34151.32     57750       474     31500    270000
------------------------------------------------------------------------------

.
. * comparing MINORITY by statistics of YR_WORK
. tabstat yr_work, statistics( mean sd median count min max ) by(minority)

Summary for variables: yr_work
Group variable: minority (Employee's minority status)

      minority |      Mean        SD       p50         N       Min       Max
---------------+--------------------------------------------------------------
"Non-minority" |  17.15405  10.08181        17       370         0        35
    "Minority" |  15.95192  9.978014      14.5       104         0        34
---------------+--------------------------------------------------------------
         Total |   16.8903  10.06094        17       474         0        35
------------------------------------------------------------------------------

.
. * comparing MINORITY by statistics of HIGHEST education achieved
. tabstat highest, statistics( mean sd median count min max ) by(minority)

Summary for variables: highest
Group variable: minority (Employee's minority status)

      minority |      Mean        SD       p50         N       Min       Max
---------------+--------------------------------------------------------------
"Non-minority" |  13.69459  2.942304        15       370         8        21
    "Minority" |  12.76923  2.555142        12       104         8        19
---------------+--------------------------------------------------------------
         Total |  13.49156  2.884846        12       474         8        21
------------------------------------------------------------------------------

.
. *---------------------------------------------------------------
. *-----Are salaries for minority employees lower than those for majority employees?--------
. *---------------------------------------------------------------
.
. * use a two sample hypothesis test, use z distribution N > 100
. * ttest for MINORITY (2 categories grouped) and ANNUAL_SAL (continuous)
. * is mean salary for minority and non-minority workers the same or non-minority is larger?
. * alpha 0.05, p= 0.0001 < 0.1, reject null, mean non-minority salary is larger than minority
```

```
. ttest annual_sal, by(minority)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
"Non-min |     370    72046.62    1876.136    36088.19    68357.36    75735.88
"Minorit |     104    57427.88    2239.967    22843.28    52985.44    61870.33
---------+--------------------------------------------------------------------
Combined |     474    68839.14    1568.622    34151.32     65756.8    71921.47
---------+--------------------------------------------------------------------
    diff |            14618.74    3734.222                7280.981    21956.49
------------------------------------------------------------------------------
    diff = mean("Non-min) - mean("Minorit)                        t =   3.9148
H0: diff = 0                                     Degrees of freedom =      472

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9999         Pr(|T| > |t|) = 0.0001          Pr(T > t) = 0.0001

.
. *---------------------------------------------------------------
. *---------------What characteristics are associated with salary?--------------------------
. *---------------------------------------------------------------
.
. * use multivariate regression/correlation between ANNUAL_SAL (continuous) and other variables
(grouped and continuous)
.
. * create dummy variables for POSITION since it has 3 categories
. * clerical = 1, custodial = 2, manager = 3
. * dummy variable MANAGER, "manager" = 1, "custodial" and "clerical" = 0
. generate manager = 1 if position == 3
(390 missing values generated)

. replace manager = 0 if position != 3
(390 real changes made)

.
. * dummy variable CLERICAL, "clerical" = 1, "custodial" and "manager" = 0
. generate clerical = 1 if position == 1
(111 missing values generated)

. replace clerical = 0 if position != 1
(111 real changes made)

.
. * multivariate regression for ANNUAL_SAL using all variables with POSITION replaced by dummies
. regress annual_sal yr_work minority sex highest manager clerical

      Source |       SS           df       MS      Number of obs   =       474
-------------+----------------------------------   F(6, 467)       =    210.47
       Model |  4.0273e+11         6  6.7122e+10   Prob > F        =    0.0000
    Residual |  1.4894e+11       467   318919389   R-squared       =    0.7300
-------------+----------------------------------   Adj R-squared   =    0.7266
       Total |  5.5167e+11       473  1.1663e+09   Root MSE        =     17858

------------------------------------------------------------------------------
  annual_sal | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     yr_work |  -212.1985   81.97024    -2.59   0.010    -373.2746   -51.12228
    minority |   -4388.26   2054.785    -2.14   0.033    -8426.029    -350.491
         sex |  -10965.61   1918.129    -5.72   0.000    -14734.85   -7196.381
     highest |   2916.881   392.0631     7.44   0.000     2146.455    3687.308
     manager |   44957.17   4937.237     9.11   0.000     35255.22    54659.12
    clerical |  -8759.841   3975.592    -2.20   0.028    -16572.11   -947.5763
       _cons |   37771.15    5597.02     6.75   0.000     26772.69    48769.61
------------------------------------------------------------------------------

.
. * correlation matrix for regression of ANNUAL_SAL
. pwcorr annual_sal yr_work minority sex highest manager clerical, obs sig star (5)
```

```
             | annual~l  yr_work minority      sex  highest  manager clerical
-------------+------------------------------------------------------------------
  annual_sal |   1.0000
             |
             |      474
             |
     yr_work |  -0.0841   1.0000
             |   0.0674
             |      474      474
             |
    minority |  -0.1773* -0.0495   1.0000
             |   0.0001   0.2821
             |      474      474      474
             |
         sex |  -0.4499*  0.0665  -0.0757   1.0000
             |   0.0000   0.1485   0.0999
             |      474      474      474      474
             |
     highest |   0.6606* -0.0474  -0.1329* -0.3560*  1.0000
             |   0.0000   0.3033   0.0038   0.0000
             |      474      474      474      474      474
             |
     manager |   0.8042* -0.0021  -0.1926* -0.3137*  0.6053*  1.0000
             |   0.0000   0.9640   0.0000   0.0000   0.0000
             |      474      474      474      474      474      474
             |
    clerical |  -0.6977*  0.0078   0.0885   0.4059* -0.3914* -0.8393*  1.0000
             |   0.0000   0.8648   0.0541   0.0000   0.0000   0.0000
             |      474      474      474      474      474      474      474
             |
```

```
.
. * YR_WORK, MINORITY, SEX, HIGHEST, CLERICAL are all associated with ANNUAL_SAL
.
. *-----------------------------------------------------------------
. /*-----Do minority and majority employees differ with respect to the characteristics that
> *are associated with salary?----------------- */
. *-----------------------------------------------------------------
.
. * differ = ttest or anova (one continuous)
. * association = chi square (two grouped)
.
. /* use various statistical tests comparing means of minority and non-minority categories of
> MINORITY with other variables*/
.
. * determine significance between MINORITY (grouped 2 category) and YR_WORK (continuous)
. * use 2 sample hypothesis test, z distibution
. * p > alpha for two tailed test so accept null hypothesis, no significant difference
. ttest yr_work, by(minority)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
"Non-min |     370    17.15405    .5241283    10.08181     16.1234    18.18471
"Minorit |     104    15.95192    .9784248    9.978014    14.01145     17.8924
---------+--------------------------------------------------------------------
Combined |     474     16.8903    .4621145    10.06094    15.98224    17.79835
---------+--------------------------------------------------------------------
    diff |            1.202131    1.116445               -.991686    3.395948
------------------------------------------------------------------------------
    diff = mean("Non-min) - mean("Minorit)                        t =   1.0767
H0: diff = 0                                     Degrees of freedom =      472

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8589         Pr(|T| > |t|) = 0.2821          Pr(T > t) = 0.1411

.
. * determine significance between MINORITY (grouped 2 category) and POSITION (grouped 3
category)
. *p < alpha so associated
```

```
. tabulate minority position, chi2 column expected

+--------------------+
| Key                |
|--------------------|
|     frequency      |
| expected frequency |
| column percentage  |
+--------------------+

    Employee's |
      minority | Employee's position at Safecorp
        status |  Clerical  Custodial    Manager |     Total
---------------+---------------------------------+----------
"Non-minority" |       276         14         80 |       370
               |     283.4       21.1       65.6 |     370.0
               |     76.03      51.85      95.24 |     78.06
---------------+---------------------------------+----------
    "Minority" |        87         13          4 |       104
               |      79.6        5.9       18.4 |     104.0
               |     23.97      48.15       4.76 |     21.94
---------------+---------------------------------+----------
         Total |       363         27         84 |       474
               |     363.0       27.0       84.0 |     474.0
               |    100.00     100.00     100.00 |    100.00

          Pearson chi2(2) =  26.1718   Pr = 0.000

.
. * determine significance between MINORITY (grouped 2 category) and SEX (grouped 2 category)
. * p > alpha so not associated
. tabulate minority sex, chi2 column expected

+--------------------+
| Key                |
|--------------------|
|     frequency      |
| expected frequency |
| column percentage  |
+--------------------+

    Employee's |
      minority |    Employee's sex
        status |    "Male"     Female |     Total
---------------+----------------------+----------
"Non-minority" |       194        176 |       370
               |     201.4      168.6 |     370.0
               |     75.19      81.48 |     78.06
---------------+----------------------+----------
    "Minority" |        64         40 |       104
               |      56.6       47.4 |     104.0
               |     24.81      18.52 |     21.94
---------------+----------------------+----------
         Total |       258        216 |       474
               |     258.0      216.0 |     474.0
               |    100.00     100.00 |    100.00

          Pearson chi2(1) =   2.7139   Pr = 0.099

.
. * determine significance between MINORITY (grouped 2 category) and HIGHEST (continuous)
. * use 2 sample hypothesis test, z distibution
. * p < alpha for two tailed, so reject null, there is significant difference
. ttest highest, by(minority)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
"Non-min |     370    13.69459    .1529631    2.942304    13.39381    13.99538
"Minorit |     104    12.76923    .2505523    2.555142    12.27232    13.26614
```

```
---------+--------------------------------------------------------------------
Combined |     474    13.49156    .1325054    2.884846    13.23119    13.75193
---------+--------------------------------------------------------------------
    diff |            .9253638    .3176764                .3011288    1.549599
------------------------------------------------------------------------------
    diff = mean("Non-min) - mean("Minorit)                        t =   2.9129
H0: diff = 0                                     Degrees of freedom =      472

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9981         Pr(|T| > |t|) = 0.0038          Pr(T > t) = 0.0019

. 
. *-----------------------------------------------------------------
. /*---Do differences in characteristics of minority and majority employees "explain"
> the lower salaries of minority employees?------------- */
. *-----------------------------------------------------------------
. 
. * No/Not necessarily, correlation does not mean causation
. 
```