

## **Introduction**

A moderately-sized company located in the Pacific Northwest states that it gives equal opportunities for all of its employees, regardless of gender, race or any other categorization. The company is going to evaluate the employees' job rating and promote some outstanding employees. The company is interested in the association between personality type and job ratings.

However, though the company claims that it values diversity and gives equal opportunities for all employees, there might exist other possible relationships that could influence job rating and promotion, such as gender and race.

The goal of this study is to find out what sort of people get better job ratings and finally get promoted. The study is focus on examine what kind of personality type is favored in the job rating and promotion process. Besides, the study is also going to examine if the company favor employees based on gender and race.

## **Methods**

### **Datas**

The data set includes 437 samples of employees from Pacific Northwest company. The employees in this data set were with the company for at least two years (one full year beyond the probationary period), and this subset of the data includes only employees classified at the supervisory level or below (no managers or executives) at the time of the most recent evaluation cycle. Salary, gender, race and personality are the 4 independent variables that potentially predict job-rating and promotion: salary is 12-month adjusted FT (in USD), which is transformed to

algorithm form to satisfy the assumption of normality for convenience; gender is the dummy variable for gender identification (coded 0 for male and 1 for female); and race is a categorical variable (coded 0 for white, 1 for Asian, and 2 for other). Personality is measured by 11 items on the self-rated instrument asked how well the following word/phrase describes the employee on a 1 (not at all like me) to 5 (very much like me) scale, including “do a thorough job”, “original”, “reserved”, “curious”, “reliable”, “imaginative”, “quite”, “shy”, “inventive”, “perseveres” and “sticks to a plan”. Job-rating is a score assigned by the employee’s direct supervisor on a 0–100 scale. Promotion is a dummy variable indicating if employee was promoted within the past 11-months (coded 1 for yes and 0 for no). The dataset has no missing data in this final version of the data set.

The collection of data is conducted by Pacific Northwest company.

### **Statistics Techniques**

I conducted a principle component analysis with promax rotation to find out how many factors I need to predict personality, and conducted exploratory factor analysis on the 11 personality variable to develop summary measures of personality. There are 11 item in our data, which are potentially related to personality. Eigen decomposition was used to proportionally transform covariance matrix of data. Promax rotation is sort of rotation that includes correlation among factors. A certain number of eigenvalue was selected to do the reduction process.

I constructed correlation tests to assess multicollinearity. I conducted t tests to assess the correlation between gender and others; I conducted ANOVA tests for race and others to see if there are any significant correlation exists with 95% confidence level. Since race and gender are

both categorical variables, I made a contingency table and then conducted a chi-square independent test the correlation between them.

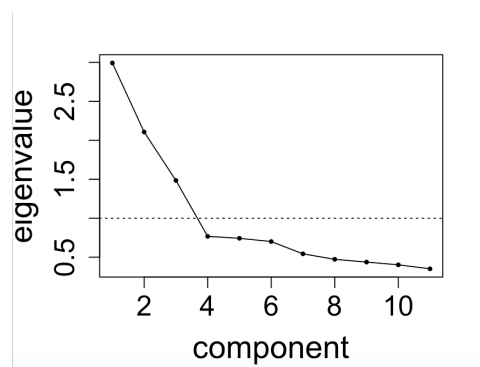
We are interested in the relationships between dependent variables (job-rating and promotion) and independent variables (gender, race, log\_salary and personality factors).

Bivariate regression was conducted to assess the relationships between dependent variables and independent variables. The coefficients are compared to coefficients in regressions with controls of possible confounding variables. Confounding effects on gender and race is specifically examined in this study.

Backward elimination was used in model selection. After running backward elimination, multiple linear regression was conducted (using factors chosen in backward elimination) to estimate job rating, and logistic regression was conducted to estimate promote. ANOVA and coefficient of determination was used to assess goodness of fit.

AUC and Hosmer-Lemeshaw test was conducted to see if the predicted probability of being promoted by the model aligns with the preserved probability from the dataset.

## Results



*Figure 1: The relationship between eigenvalue and component*

Figure 1 indicates that all the variance from variables accounted by the components after the third is not statistically significant. Around 70% of the variance is preserved from the original sample. This study groups the 11 item samples to 3 factor variables.

Loadings:			
	Factor1	Factor2	Factor3
thorough	0.76600000000000014210855		
original			0.50100000000000000888178
reserved		0.776000000000000023092639	
curious			0.461000000000000020872193
reliable	0.75900000000000007993606		
imaginative			0.42199999999999986233234
quiet		0.83499999999999964472863	
inventive			0.84299999999999971578291
perserveres	0.70699999999999961808328		
shy		0.68799999999999944932938	
stickplan	0.537000000000000032862602		
	Factor1	Factor2	Factor3
SS loadings	2.0470000000000001527667	1.8040000000000000479616	1.377000000000000017764
Proportion Var	0.18599999999999986677	0.164000000000000068834	0.125000000000000000000
Cumulative Var	0.18599999999999986677	0.349999999999999777955	0.474999999999999777955
Factor Correlations:			
	Factor1	Factor2	Factor3
Factor1	1.000	-0.143	0.297
Factor2	-0.143	1.000	-0.397
Factor3	0.297	-0.397	1.000
Test of the hypothesis that 3 factors are sufficient.			
The chi square statistic is 46.109999999999943157 on 25 degrees of freedom.			
The p-value is 0.0062300000000000294487			

*Table1: Factors loading and Correlation between 3 factors*

According to table1, we can see that there is no weak loading or cross loading. Table 1 indicates that factor 1 is correlated with thorough, reliability, persevere and stick to plan; factor 2 is correlated with reserved, quiet and shy; factor 3 is correlated with original, curious, imaginative and inventive. From factor correlation of table1, we can see that the correlation between factor 1 and factor 2 is -0.143, correlation between factor 2 and factor 3 is -0.397, and the correlation between factor 1 and factor 3 is 0.297. All the 3 correlations are moderate.

	gender	race	log_salary	personality1	personality2	personality3
gender	—					
race	0.80906	—				
log_salary	0.000251	0.89215	—			
personality1	0.000000	0.57346	0.28311	—		
personality2	0.037375	0.36553	0.56559	0.34849	—	
personality3	0.033361	0.128	0.18065	0.00000	0.000008	—

*Table2: p-values of correlation tests*

(For better understanding, I named factor1 personality1, factor2 personality2 and factor3 personality3)

Table 2 is the summary of p-values of the correlation tests between independent variables. From table 2 we can see that the factor gender is significantly correlated with log salary, personality1, personality2 and personality3, and there exist significant correlation between personality1 and personality3, personality2 and personality3. Only 1 confounding effect was detected after running the tests. The coefficient of bivariate regression of job rating and gender is -1.394056, but after controlling for the factor personality1, the coefficient changed to -0.602998, which significantly increased. Therefore, personality1 is the potential confounding variable for gender.

Backward elimination based on AIC for Multiple classical linear model chooses Asian, others, log\_salary, personality1, personality2 and personality3 as predictors. For the logistic regression of promote, backward step-wise model selection based on AIC chooses gender, other, salary, consistency and job rating.

```

Call:
lm(formula = jobrating ~ asian + other + log_sal + Factor1 +
    Factor2 + Factor3, data = new_data)

Residuals:
    Min       1Q   Median       3Q      Max
-16.21119180681425575585 -2.99690743027208883476 -0.04087968088704990732  3.01097993965130239502
 14.96013568103352930905

Coefficients:
            Estimate      Std. Error    t value      Pr(>|t|)
(Intercept) 58.9899638939776451707  5.8997525506490093861  9.998720000000000496 < 0.000000000000000222 ***
asian       -0.8723937530330221346  0.5087476450766659219 -1.714790000000000036  0.087105 .
other       -1.5883243884181597050  0.6361456560566348672 -2.496789999999999843  0.012905 *
log_sal      2.9383945547136818277  1.1540717878155928311  2.546110000000000095  0.011241 *
Factor1      2.1701253199686916240  0.2543051827151115996  8.533549999999999969 0.00000000000000024578 ***
Factor2     -0.5052673064315249540  0.2506921712848695361 -2.015489999999999782  0.044475 *
Factor3      1.3824624802175682348  0.2680078046574181916  5.158290000000000042 0.00000038117977785486 ***
---
Signif. codes:
  0 '***' 0.0010000000000000020817 '***' 0.0100000000000000020817 '*' 0.05000000000000000277556 '.'
 0.1000000000000000055511 ' ' 1

Residual standard error: 4.727281703526297285123 on 430 degrees of freedom
Multiple R-squared:  0.1719638940444692676,    Adjusted R-squared:  0.1604099018683454725
F-statistic: 14.88350445656632282 on 6 and 430 DF,  p-value: 0.00000000000001740631739760767012

```

*Table3: Multiple linear regression*

Based on the backward elimination steps, asian, others, log\_salary, personality1, personality2 and personality3 are chosen as predictors to run the multiple regression linear test. According to table3, other, log\_salary, personality1, personality2 and personality3 are significant. Only Asian is not significant at this level. The intercept is the reference group (white male), and this value is 58.99, which is significant with p value 0.00000. Average job rating decrease by 0.87239 when the race is Asian, which is not significant with p value 0.087105. Average job rating would decrease by 1.588324 when the race is other, which is significant with p value 0.012905. A 1% increase in salary would lead to a increase of 2.9384 in job rating, which is significant with p value 0.011241. 1 score in any personality type in personality1 would increase job rating by 2.170125, which is significant with p value 0.0000. a score in any personality type in personality2 would increase job rating by 0.505267, which is significant with p value 0.044475. 1 score in any personality type in personality3 would increase job rating by 1.38246, which is significant with p value 0.0000.

According to table3, the coefficient of determination in the multiple linear regression model is 0.172, indicating that 17.2% of variability of job rating could be explained by the model. The F-statistic is 14.88 with degrees of freedom 6 and 430.

```
Call:
glm(formula = promote ~ gend_F + other + log_sal + Factor1 +
    jobrating, family = binomial(link = "logit"), data = new_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4974820718262136232  -0.9050162406085718869  -0.6407617724320378017   1.1881598081001540912
 2.1217262498695692585

Coefficients:
              Estimate              Std. Error      z value    Pr(>|z|)
(Intercept) -10.74240906157987218705    3.18947518632018356399  -3.36807999999999963  0.00075694 ***
gend_F       -0.61655899792084978284    0.22528550093242696861  -2.736790000000000056  0.00620420 **
other        -0.79662513640637711010    0.31009752971104653785  -2.568950000000000067  0.01020071 *
log_sal      1.33232994342172839097    0.56924657439858761254   2.340510000000000090  0.01925718 *
Factor1      0.34621351345972367630    0.13331346187370415879   2.59698999999999910  0.00940451 **
jobrating    0.04929278518785999708    0.02251039172132804703   2.189779999999999838  0.02854026 *
---
Signif. codes:
  0 '***' 0.00100000000000000020817 '***' 0.01000000000000000020817 '**' 0.050000000000000000277556 '.'
  0.10000000000000000055511 '.' 1

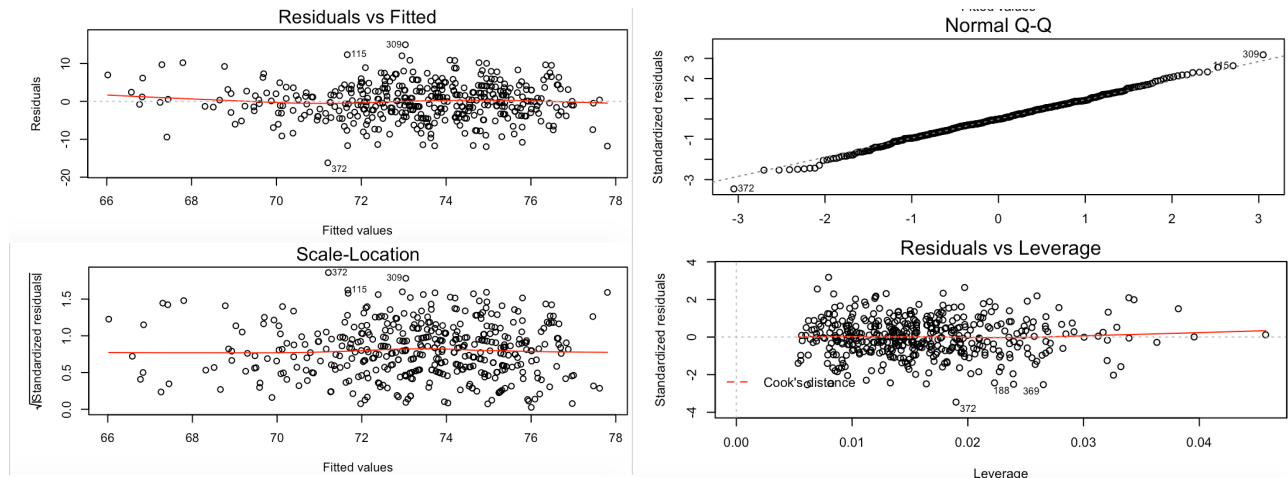
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 556.77433748090288645 on 436 degrees of freedom
Residual deviance: 512.15829159541328863 on 431 degrees of freedom
AIC: 524.15829159541328863

Number of Fisher Scoring iterations: 4
```

#### Table4: logistic regression

According to table4, all variables are significant. The estimate of intercept indicates that the odds of White and Asian male passing the promotion is 0.0000216 with p value 0.000757. Gender and race are the only 2 dichotomous variables in the list, so we only need to focus on the odds ratio of gender and race in this analysis. The log odds ratio of gender is -0.61656 with p value 0.0062042, and it suggests that the odds of employees passing the promotion is decreasing by 0.61656 if the employee is female. The log odds ratio of other(race) is -0.796625 with p value 0.0102007, and it suggests that the odds of getting promotion is decreased by 0.796625 if the employee is other race.



*Figure2: residual plots of multiple regression model*

According to the Residuals vs. Fitted plot, there are no significantly curved pattern.

According to the normal QQ plot, we can see the normality of residual's distribution.

After running the Hosmer-Lemeshaw test, we can see that the Chi-squared test statistic is 9.696 with p-value 0.287 and 8 degree of freedom. Since the p value is very large, there is no sufficient evidence to reject the null hypothesis that the expected probability is the same with the observed probability across all subgroups.

## Conclusion

It can be concluded from the study that the race White and Asian, personality related to personality1 and personality3 (thorough, reliability, persevere, stick to plan, original, curious, imaginative and inventive) can get a better score in job ratings. Employees with personality2 (reserved, quiet and shy) tend to get lower scores in job rating. Compared to Asian and White, other race is significantly graded lower in job rating. Higher salary is related to significantly higher job ratings.



Compared to males, females are significantly less favored in promotion process.

Compared to Asian and White, other race is significantly less favored in promotion. Higher salary is related to higher chance of passing the promotion. Having qualities like thorough, reliability, persevere and stick to plan is significantly increase chance of passing the promotion. Higher job rating score can also predict better chance of passing promotion.