**Data Analysis Report on Job Rating and Promoting Employees**

**Course: MA 684**

**Author: Zhe Zhao**

**Abstract**

Companies value diversity and equal opportunity more and more in today’s society. However, it is common that there is still implicit difference in treating employees with different backgrounds and personalities. The purpose of this Data analysis report is to help Pacific Northwest Company to determine if there is any different treatment based on gender and race, and to examine what personality would the company prefer when evaluating the performance of employees and deciding who is to promote. This report finds that gender and race difference in job rating and odds of getting promoted still exists. The company would favor consistency and creativity in employees’ personalities.

**Introduction**

Pacific Northwest Company is an equal-opportunity employer with employees from different backgrounds and with different personalities. It is important for the Human Resource Department to evaluate employees’ job rating and promote certain amount of people in order to, firstly, self-reflecting and find problems to improve efficiency and secondly, ensure the company is run based on fair and equal standards to every employees.

However, besides the actual performance of employees, job rating and the chance of getting promoted may correlated with other characteristics such as gender, race, salary level and personalities. To ensure that the company has no discrimination against minority group or any gender, the data analysis is conducted to examine the effects of employees’ demographic features on their job rating the chance of getting promoted. The analysis will also focus on helping HR department to determine what personality is favored by the company and what kind of person the company should hire.

The goal of the study is to examine what kind of employees would get a higher job rating and get promoted. Is the company favor people based on gender and race? What kind of personality is aligned with company’s standards? This report does not intend to make any claim about Pacific Northwest’s internal policy and would not provide strategy advice to the decision-making of HR department.

**Methods**

**Details of Data**

This report collects data of job rating from 437 employees comprising 218 female and 219 men. All employees are classified at or below the supervisory level with at least two years’ experience in the company. There are 4 variables considered as potential predictors of job ratings, salary, gender, race and personality. Salary is 12-month adjusted full-time salary in US dollar. For the convenience of this study, salary is transformed to algorithm form to satisfy the assumption of normality. Gender is coded with 0 as male and 1 as female. Race has three levels: white coded as 0, Asian coded as 1 and others coded as 2. In this study, we use white as a baseline level for the comparison and race, therefore, is divided into two factor variables, Asian (0 as white or others and 1 as Asian) and other ((0 as white or Asian and 1 as others)). Personality is measured by a 5 points scale (from 0 “not at all like me” to 5 “very much like me”) self-rating questionnaire based on 11 items: “do a thorough job”, “original”, “reserved”, “curious”, “reliable”, “imaginative”, “quite”, “shy”, “inventive”, “perseveres” and “sticks to a plan”.

Another dependent variable of interests is promote indicating if the employee was promoted in the last 11 months. The variable promote is coded with 1 as yes and 0 as no.

The process of data collection is conducted by researchers of Pacific Northwest Company.

**Statistic Techniques**

It is very common to conduct principle component analysis for the data collected by self-reporting 5 points scale survey. The purpose of the analysis is to reduce the column dimension of variable matrix while keeping as much information as possible. In this study, we have 11 items that are potentially related to the personality and we want to reveal latent factors of observed values. Eigen decomposition to the proportional transformed covariance matrix of sample data is applied. Here, transformation (or projection) is defined by Promax rotation, which is a form of Oblique rotation that allows correlation exists among factors. Eigenvalues are often referred to the diagonal of the transformed covariance matrix (a measurement of variance from all variables accounted by this factor). The reduction process is conducted by choosing a number of eigenvalues so that firstly, errors between the original dataset and new chosen dataset could be minimized, and secondly, the preserved variance (sum of eigenvalues) of the original dataset could be maximized.

Multicollinearity, confounding variable would diminish the validation of linear modeling. For multicollinearity, we construct correlation test (Welch two samples t-test between gender and others, Chi-squared tests among log-salary and factors gained from principal component analysis) among independent variables and ANOVA test between race and other variables to check if any significant correlation exists with 95% confidence level. For confounding variables, bivariate regressions between each independent variable and dependent variables are fitted, and coefficients are compared to coefficients in regressions with controlling for the possible confounding variables. This study particularly examines the confounding effects on gender and race.

For the estimation of job ratings, multiple linear regression is fitted and followed by backward step-wise model selection based on AIC and BIC in order to determine parsimonious model. Coefficient of determination and ANOVA table are main tools to check the goodness of fit. The model validation includes checking the normality of residuals and the detection of leverage points.

For the estimation of promote, this study chooses logistic regression due to the presence of binary data. The linear link function is defined by logit function, the log odds ratio of probability of being promote in the last 11 months against probability of not being promoted. The model selection is based backward step-wise with AIC and BIC as criterions. Analysis of deviance table with Chi-squared test is used to check goodness of fit as well as Hosmer-Lemeshaw test to check if the model assumption aligns with the observed probability. For model assessment, this study plots deviance residuals to examine the linearity between the link function and fitted values, and calculates the dispersion from model’s Pearson residuals to check over-dispersion

**Results**

Eigenvalues plots against component indicates (figure 1) that after the third component, the variance from all variables accounted by this component is not significant. By three components, about 70% of variance is preserved from the original sample. This study therefore reduces the original 11 item sample to three factor scores variables. Factor loadings in table 1 indicates no weak loading or cross loading. Based on table 1, three factors are summarized into perspectives of the following: factor 1 as “consistency” due to its correlation with thorough, reliability, persevere and stick to plan; factor 2 as “social difficulty” due to its correlation with reserved, quiet and shy; factor 3 as “creativity” due to its correlation with original, curious, imaginative and inventive. We examine the correlation table of three factors and find that the correlation between factor 1 and factor 2 is -0.143, correlation between factor 1 and factor 3 is 0.297 and correlation between factor 2 and factor 3 is -0.397. Three correlations are moderate.

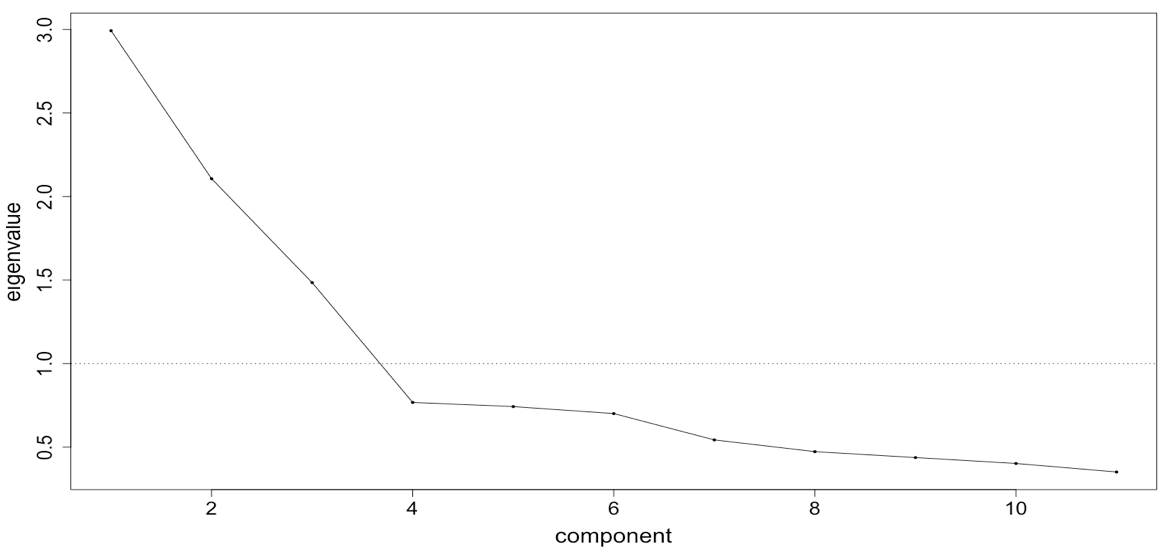


Figure 1 eigenvalue and components plot

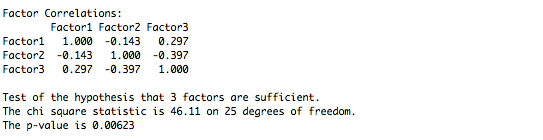


Table 2 Correlation table of three factors

Table 1 factor loading

Table 3 is the summary of p-value of correlation test (95% confidence level) among independent variables. The null hypothesis of the test states that no significant correlation between two variables. We find that variable gender has significant correlation with log-salary, consistency, social difficulty and creativity. Besides above, there are significant correlation between creativity and consistency and between social difficulty and creativity. For the detection of confounding variable, we only find one potential confounding effect in linear modeling. The coefficient of the bivariate regression of job rating against gender was -1.3941. After controlling for consistency, the coefficient of gender increases by almost 40% to -0.6030. Consistency is a potential confounding variable to gender. In addition, we also fit each interaction terms into the bivariate regression with job rating. There is no significant interaction effect detected.

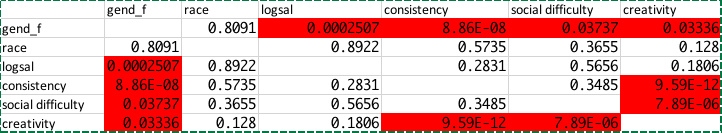


Table 3 p-value of correlation tests table

Backward step-wise model selection based on AIC for Multiple classical linear model chooses Asian, other, log salary, consistency, social difficulty and creativity as predictors; BIC criterion chooses less predictors: log salary, consistency and creativity. We conduct F-test to see if model with more variables could significantly increase the goodness of fit and the p-value is 0.01767, which suggests that model selected by AIC significantly improve the prediction power compared to model selected by BIC. For the logistic regression of promote, backward step-wise model selection based on AIC chooses gender, other, salary, consistency and job rating. Backward step-wise model selection based on BIC chooses gender, others, log-salary and consistency. We compared two models by conducting analysis of deviance Chi-squared test and the p-value is 0.02725, which also suggests that model selected by AIC significantly improves the prediction power compared to model selected by BIC in logistic case.

According to the summary table (table 4) of the selected multiple linear regression, other race, log-salary, consistency (factor 1), social difficulty (factor 2) and creativity (factor 3) are significant at 95% level except for Asian. The intercept is the coefficient of baseline race, white, means in average, the job rating of whites is 58.99, and this value is significant with p-value near 0.000. The average job rating would decrease by 0.8724, which has no significant difference to whites with p-value 0.087, and by 1.5883 compared to whites if race is others with p-value 0.0129. One percent increase in salary would bring the job rating up by 1.2761 points in average with p-value 0.0112. For consistency, if the individual has one score on consistency-related personality, the job-rating would increase by 2.17 in average with p-value near 0.0000. One more score on social difficulty-related personality would bring 0.5052 decrease to the average job rating with p-value 0.0445, and one more score on creativity-related personality would bring 1.3825 increase to the average job rating with p-value near 0.0000.

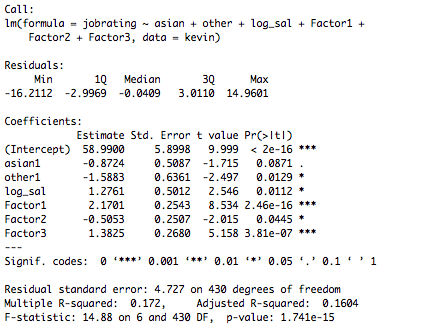
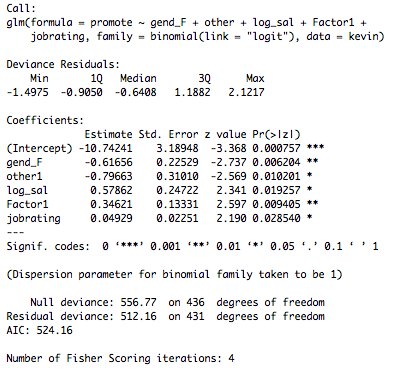
According to the summary table (table 5) of the chosen logistic linear model, all variables: gender, other race, log-salary, consistency (factor 1) and job rating, are significant at 95% level. The intercept means that the odds of race whites and Asian men getting promoted is 2.16 \* 10-5 with p-value 0.000757. The log odds ratio of gender is -0.61656, which means that the odds of female getting promoted is 53.9798% of men with p-value 0.006204. The log odds ratio of other race is -0.79663, which means that the odds of other race getting promoted is 45.085% of whites and Asian with p-value 0.010201. Log odds ratio of log-salary is 1.33233, which means that one percent increase of salary could increase 78.3575% odds to get promoted with p-value 0.01926. The log odds ratio of consistency is 0.34621, which means one unit increate in consistency scores would increase 41.37% odds of getting promoted with p-value 0.009405. The log odds ratio of job rating is 0.04929, which means one more score on job rating would have 5.053% increase of odds of getting promoted with p-value of 0.02854.

Table 5 Logistic regression

Table 4 Multiple linear regression

**Model Validation**

The coefficient of determination in multiple linear model is 0.172. 17.2% of variability of job rating could be explained by the model. The F-statistic of the model is 14.88 with degrees of freedom 6 and 430. Since the p-value is nearly zero, we reject the null hypothesis that the model has no improvement from the null model.

From the residual plots against fitted value in figure 2, we find no significant curved pattern or non-constant variance. The normal QQ plot shows indicates the normality of residuals’ distribution. However, we find point 372 could be a potential bad leverage point, which means the point is away from the cluster of majority sample data and has significant effects on regression fitting. Point 372 is indicates a white female with salary 237950, which is higher than the third quartile, and her job rating is 55, the minimum of whole dataset. Since it is a random observation and not an error, we decide to keep the data.

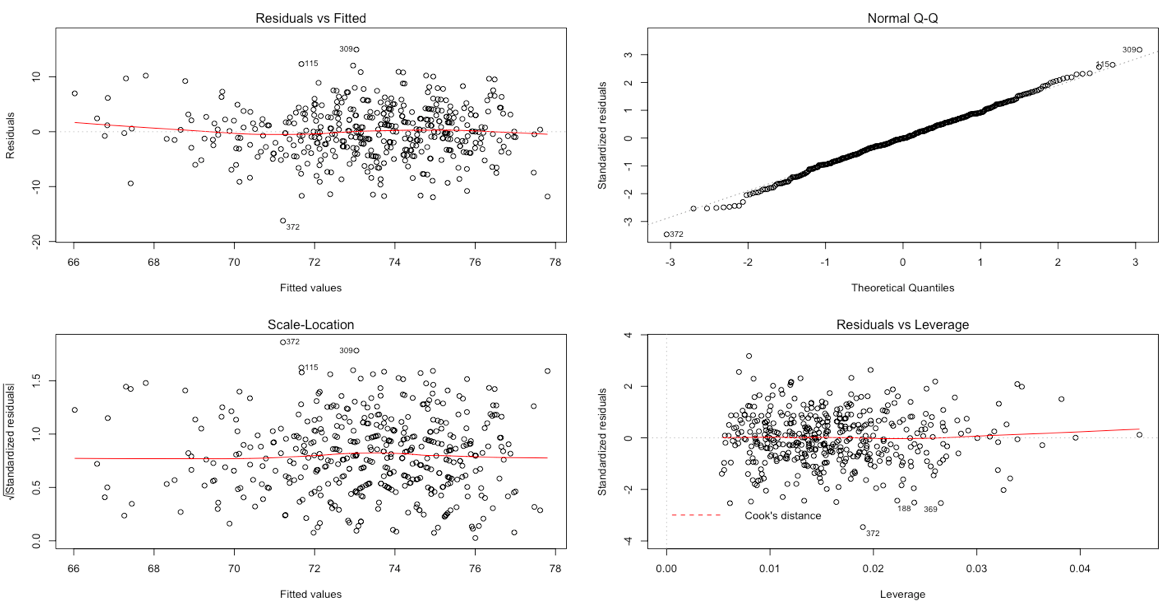


Figure 2 diagnostic plots

The Analysis of deviance table conducts Chi-squared test when each variable is added into model to check if the variable significantly improves the model. The p-value for the test of gender is 0.000612, for other race is 0.0063958, for log salary is 0.0030921, for consistency is 0.005877 and for job rating is 0.0272509. Above p-value suggests that we could reject the null hypothesis that each variable equal zero and conclude that each variable significantly improve the model.

Hosmer-Lemeshaw test was conducted to check if the predicted probability of being promoted by the model aligns with the overserved probability from the dataset. The model population is divided into 10 equally subgroups and the Chi-squared test statistic is 9.696 with degrees of freedom 8 and the p-value is 0.287. We failed to reject the null hypothesis, which states that expected probability is the same with the observed probability across all subgroups.

We expect the LOWESS (locally weighted smoothing) fit line in deviance residuals against fitted value plot to be a horizontal line near zero since the assumed distribution of each observation is a Bernoulli trial. The horizontal line suggests there is no other type of link or higher order of polynomial variables needed in the model. As we can see from figure 3, deviance residuals are distributed as two exponential decaying curves with the mean nearly to be zero. The LOWESS fit curve shows a little curved pattern at the start of the line which suggests there might be underlying higher order of polynomial variables or wrong link function. However, since the curved pattern is not visually significant, it is sufficient to assume deviance residuals has the same distribution as expected.

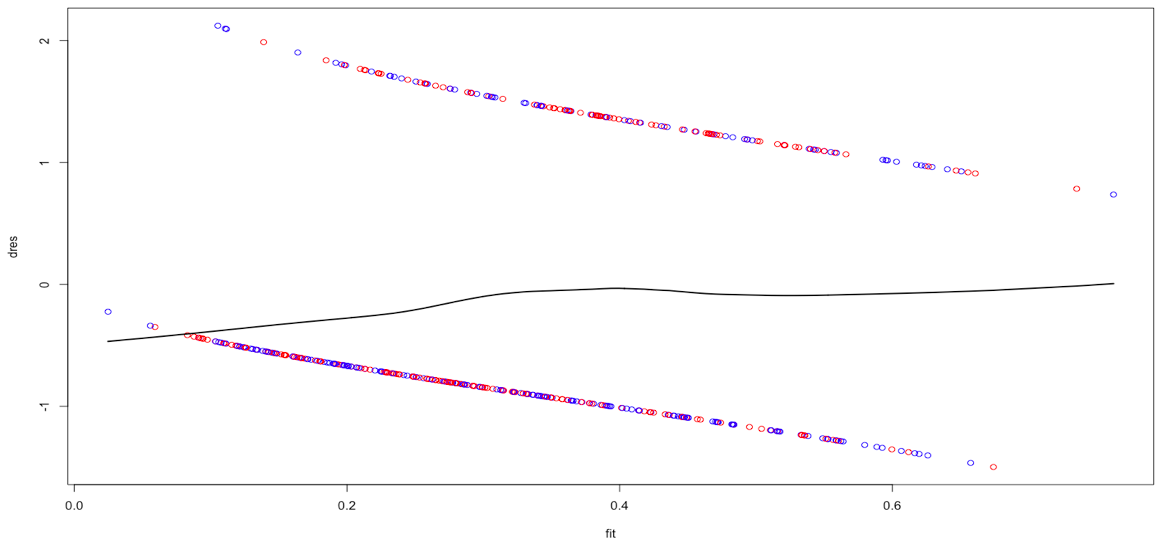


Figure 3 deviance residuals plots

The estimated dispersion of the model is calculated by the sum of squared Pearson residuals divided by degrees of freedom. The result is 0.994, which is very close to the expected dispersion of binomial distribution, 1. When we re-test the significance of each variable in model while adjusting the model dispersion as 0.994, we find the significance of variable has no change. We conclude there is no significant over-dispersion for the model fit.

**Discussion**

It is unclear about the interpretation of factor scores in the dataset. What does “unit change” means for factor score and what does the factor score measure? The study summarizes three factors as “consistency”, “social difficulty” and “creativity” base on their significant loadings to questionnaire items. Factor scores were treated as an evaluation of how much a person could related to this latent variable. Therefore, there is no practical interpretation on the unit change in any attributes in personalities. The coefficients of three factors gained from principal component analysis in both models can only be interpreted as how would a person’s job rating (or odds of promotion) change if this person has more relative connection to this factor.

Final models still exist the issue of confounding variable and multicollinearity. In the logistic regression model, gender and consistency both are treated as significant variables. However, the existence of confounding effect might distort the coefficients and p-value in the model summary might not be reliable. Similarly, the study finds potential significant collinearity between gender and three factors: consistency, social difficulty and creativity, and collinearity between consistency and creativity, between social difficulty and creativity. However, the multiple regression linear model contains all three factors. The reported interpretation of coefficients might be inflated. The further investigation should involve more data points to be enough to illustrate the correlation among data. Potential strategy might also include bootstrapping to residuals.

The odds of white and Asian men getting promoted in the last 11 months seems too low, which is only 2.16 \* 10-5. The reason might be the variable log-salary. The full interpretation of this odds is the chance of getting promoted if the individual is white or Asian men with no increase of salary (log-salary is zero). Here, we find that the interpretation of intercept does not have practical meaning without considering the change of salary.

**Conclusion**

It can be concluded that white and Asian race with personality related to consistency and creativity would have higher job rating. Job rating with other race is significantly lower than white and Asian and further discussion is needed about the causation. Person with social difficulty inclines to get lower job rating. The increase in salary would bring an increase in the job rating significantly.

Female would have significantly less chance to get promoted compared to men. Other race has significantly less chance of promotion compared to Whites and Asian. Having more salary increase could increase the odds of promotion. Being consistency could increase odds of getting promoted. Odds of promotion would also be higher if the job rating is growing.

**Appendix**

data684 <- read.csv("final\_project\_Spr18.csv", header=T)  
  
# PCA  
prcomp(data684[,2:12], scale.=TRUE)

## Standard deviations (1, .., p=11):  
## [1] 1.7299143 1.4514530 1.2184745 0.8759282 0.8616212 0.8370284 0.7368874  
## [8] 0.6875430 0.6612250 0.6339834 0.5923436  
##   
## Rotation (n x k) = (11 x 11):  
## PC1 PC2 PC3 PC4 PC5  
## thorough -0.3392962 0.33605041 -0.2025200 0.14108382 -0.175243562  
## original -0.3863371 -0.05628313 0.2429396 -0.44852875 0.261877162  
## reserved 0.2070750 0.48058444 0.2573055 -0.19463151 -0.064850667  
## curious -0.2704026 0.04798212 0.4118869 0.49355782 -0.580641519  
## reliable -0.3680051 0.29083025 -0.1786477 0.19341711 0.070893933  
## imaginative -0.2663630 -0.04549069 0.3912965 0.43092189 0.677930706  
## quiet 0.2824091 0.46703502 0.2013804 -0.05444398 0.056305002  
## inventive -0.2374064 -0.10508703 0.5484340 -0.43796592 -0.241901834  
## perserveres -0.3507246 0.28008001 -0.2453888 -0.19872605 0.028434858  
## shy 0.2467783 0.45218938 0.1912502 0.08223439 0.179342114  
## stickplan -0.3053247 0.22782003 -0.2023731 -0.18248256 0.009218765  
## PC6 PC7 PC8 PC9 PC10  
## thorough 0.24106419 -0.08054234 0.51001994 0.51310966 -0.27662944  
## original 0.13939762 0.54186201 -0.05572148 -0.09430906 -0.38023161  
## reserved -0.07921463 -0.23208736 0.15660836 -0.35214431 -0.47650781  
## curious -0.15872209 0.17320215 -0.28324961 -0.09757330 -0.13830701  
## reliable 0.35689027 0.18312738 0.15448308 -0.54696035 0.42884739  
## imaginative -0.14918505 -0.31031748 0.06689436 0.05606145 -0.04548744  
## quiet -0.04717992 -0.07224890 0.08572365 -0.14641492 0.21976205  
## inventive 0.12555441 -0.27674105 0.16318525 0.18262666 0.45751542  
## perserveres 0.11575703 -0.45216653 -0.68389412 0.05607948 -0.05888638  
## shy 0.12100829 0.42088114 -0.29879662 0.48079061 0.21445699  
## stickplan -0.83386408 0.15100123 0.11273784 0.05380762 0.19914123  
## PC11  
## thorough 0.12567484  
## original 0.22175937  
## reserved -0.42677205  
## curious 0.10056843  
## reliable -0.19440136  
## imaginative -0.01856654  
## quiet 0.75558442  
## inventive -0.14082848  
## perserveres 0.07650677  
## shy -0.31765975  
## stickplan -0.08451118

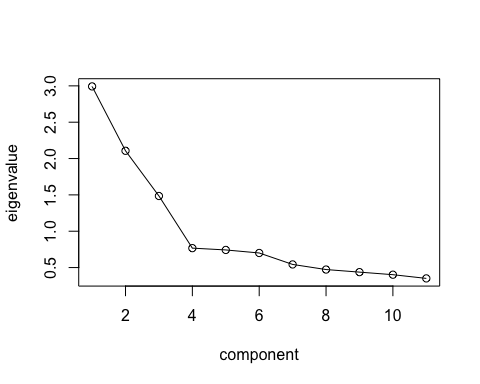
(eigens <- prcomp(data684[,2:12], scale.=TRUE)$sdev^2)

## [1] 2.9926034 2.1067158 1.4846802 0.7672503 0.7423910 0.7006166 0.5430030  
## [8] 0.4727153 0.4372186 0.4019349 0.3508709

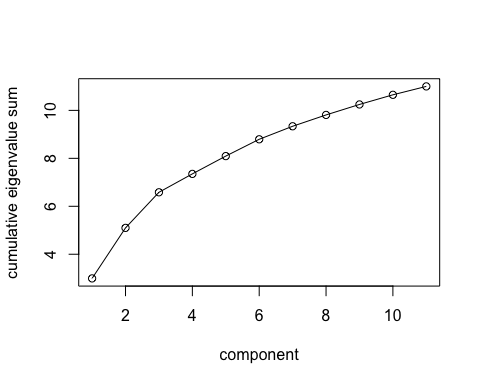
cbind(eigens,cumsum(eigens))

## eigens   
## [1,] 2.9926034 2.992603  
## [2,] 2.1067158 5.099319  
## [3,] 1.4846802 6.583999  
## [4,] 0.7672503 7.351250  
## [5,] 0.7423910 8.093641  
## [6,] 0.7006166 8.794257  
## [7,] 0.5430030 9.337260  
## [8,] 0.4727153 9.809976  
## [9,] 0.4372186 10.247194  
## [10,] 0.4019349 10.649129  
## [11,] 0.3508709 11.000000

plot(eigens,type="l",xlab="component",ylab="eigenvalue")  
points(eigens)



plot(cumsum(eigens),type="l",xlab="component",ylab="cumulative eigenvalue sum")  
points(cumsum(eigens))



print(factanal(data684[,2:12],3,rotation="promax"),cutoff=.3)

##   
## Call:  
## factanal(x = data684[, 2:12], factors = 3, rotation = "promax")  
##   
## Uniquenesses:  
## thorough original reserved curious reliable imaginative   
## 0.461 0.566 0.439 0.750 0.452 0.791   
## quiet inventive perserveres shy stickplan   
## 0.299 0.411 0.528 0.514 0.722   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## thorough 0.766   
## original 0.501   
## reserved 0.776   
## curious 0.461   
## reliable 0.759   
## imaginative 0.422   
## quiet 0.835   
## inventive 0.843   
## perserveres 0.707   
## shy 0.688   
## stickplan 0.537   
##   
## Factor1 Factor2 Factor3  
## SS loadings 2.047 1.804 1.377  
## Proportion Var 0.186 0.164 0.125  
## Cumulative Var 0.186 0.350 0.475  
##   
## 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 46.11 on 25 degrees of freedom.  
## The p-value is 0.00623

asian <- 1\*{data684$race==1}  
other <- 1\*{data684$race==2}  
fact.scores <- factanal(data684[2:12],3,rotation="promax",scores="regression")$scores  
  
kevin <- data.frame(data684[13:17],log\_sal = log(data684$salary),fact.scores, asian,other)  
kevin$race <- factor(kevin$race)  
kevin$asian <- factor(kevin$asian)  
kevin$other <- factor(kevin$other)  
cor(fact.scores)

## Factor1 Factor2 Factor3  
## Factor1 1.00000000 0.04495362 -0.3182849  
## Factor2 0.04495362 1.00000000 0.2119172  
## Factor3 -0.31828486 0.21191724 1.0000000

#muticolinearity  
t.test(log\_sal ~ gend\_F, data=kevin)

##   
## Welch Two Sample t-test  
##   
## data: log\_sal by gend\_F  
## t = -3.6922, df = 433.9, p-value = 0.0002507  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.24177991 -0.07379293  
## sample estimates:  
## mean in group 0 mean in group 1   
## 11.66447 11.82226

t.test(Factor1 ~ gend\_F,data=kevin)

##   
## Welch Two Sample t-test  
##   
## data: Factor1 by gend\_F  
## t = 5.452, df = 390.07, p-value = 8.855e-08  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.3073102 0.6539594  
## sample estimates:  
## mean in group 0 mean in group 1   
## 0.2397675 -0.2408673

t.test(Factor2 ~ gend\_F, data=kevin)

##   
## Welch Two Sample t-test  
##   
## data: Factor2 by gend\_F  
## t = 2.0881, df = 434.99, p-value = 0.03737  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.01093693 0.36143640  
## sample estimates:  
## mean in group 0 mean in group 1   
## 0.09288030 -0.09330636

t.test(Factor3 ~ gend\_F, data=kevin)

##   
## Welch Two Sample t-test  
##   
## data: Factor3 by gend\_F  
## t = 2.1345, df = 434.98, p-value = 0.03336  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.01493096 0.36212520  
## sample estimates:  
## mean in group 0 mean in group 1   
## 0.09404833 -0.09447975

chisq.test(table(kevin[,c("gend\_F","race")]))

##   
## Pearson's Chi-squared test  
##   
## data: table(kevin[, c("gend\_F", "race")])  
## X-squared = 0.42376, df = 2, p-value = 0.8091

anova(aov(log\_sal ~ factor(race), data=kevin))

## Analysis of Variance Table  
##   
## Response: log\_sal  
## Df Sum Sq Mean Sq F value Pr(>F)  
## factor(race) 2 0.047 0.023526 0.1141 0.8922  
## Residuals 434 89.446 0.206097

anova(aov(Factor1 ~ factor(race), data=kevin))

## Analysis of Variance Table  
##   
## Response: Factor1  
## Df Sum Sq Mean Sq F value Pr(>F)  
## factor(race) 2 1.01 0.50419 0.5568 0.5735  
## Residuals 434 393.01 0.90555

anova(aov(Factor2 ~ factor(race), data=kevin))

## Analysis of Variance Table  
##   
## Response: Factor2  
## Df Sum Sq Mean Sq F value Pr(>F)  
## factor(race) 2 1.77 0.88296 1.0088 0.3655  
## Residuals 434 379.88 0.87529

anova(aov(Factor3 ~ factor(race), data=kevin))

## Analysis of Variance Table  
##   
## Response: Factor3  
## Df Sum Sq Mean Sq F value Pr(>F)  
## factor(race) 2 3.53 1.76612 2.0655 0.128  
## Residuals 434 371.09 0.85506

cor(kevin[,c("log\_sal",paste0("Factor",1:3))])

## log\_sal Factor1 Factor2 Factor3  
## log\_sal 1.00000000 0.05145919 -0.02755855 -0.06415932  
## Factor1 0.05145919 1.00000000 0.04495362 -0.31828486  
## Factor2 -0.02755855 0.04495362 1.00000000 0.21191724  
## Factor3 -0.06415932 -0.31828486 0.21191724 1.00000000

cor.test(kevin$log\_sal,kevin$Factor1)

##   
## Pearson's product-moment correlation  
##   
## data: kevin$log\_sal and kevin$Factor1  
## t = 1.0747, df = 435, p-value = 0.2831  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.04255089 0.14456604  
## sample estimates:  
## cor   
## 0.05145919

cor.test(kevin$log\_sal,kevin$Factor2)

##   
## Pearson's product-moment correlation  
##   
## data: kevin$log\_sal and kevin$Factor2  
## t = -0.575, df = 435, p-value = 0.5656  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.12105030 0.06641783  
## sample estimates:  
## cor   
## -0.02755855

cor.test(kevin$log\_sal,kevin$Factor3)

##   
## Pearson's product-moment correlation  
##   
## data: kevin$log\_sal and kevin$Factor3  
## t = -1.3409, df = 435, p-value = 0.1806  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.15701899 0.02982486  
## sample estimates:  
## cor   
## -0.06415932

cor.test(kevin$Factor1,kevin$Factor2)

##   
## Pearson's product-moment correlation  
##   
## data: kevin$Factor1 and kevin$Factor2  
## t = 0.93853, df = 435, p-value = 0.3485  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.04905793 0.13817564  
## sample estimates:  
## cor   
## 0.04495362

cor.test(kevin$Factor1,kevin$Factor3)

##   
## Pearson's product-moment correlation  
##   
## data: kevin$Factor1 and kevin$Factor3  
## t = -7.0025, df = 435, p-value = 9.591e-12  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.4001426 -0.2313887  
## sample estimates:  
## cor   
## -0.3182849

cor.test(kevin$Factor2,kevin$Factor3)

##   
## Pearson's product-moment correlation  
##   
## data: kevin$Factor2 and kevin$Factor3  
## t = 4.5226, df = 435, p-value = 7.888e-06  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.1205081 0.2997630  
## sample estimates:  
## cor   
## 0.2119172

#bivariate regressions  
t.test(jobrating ~ gend\_F, data=kevin)

##   
## Welch Two Sample t-test  
##   
## data: jobrating by gend\_F  
## t = 2.8469, df = 433.69, p-value = 0.004624  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.4316255 2.3564856  
## sample estimates:  
## mean in group 0 mean in group 1   
## 74.09589 72.70183

anova(aov(jobrating ~ race, data=kevin))

## Analysis of Variance Table  
##   
## Response: jobrating  
## Df Sum Sq Mean Sq F value Pr(>F)  
## race 2 69.5 34.771 1.3082 0.2714  
## Residuals 434 11535.4 26.579

summary(lm(jobrating ~ log\_sal, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ log\_sal, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.2824 -3.4167 0.2268 3.2502 14.6816   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.1327 6.3687 8.971 <2e-16 \*\*\*  
## log\_sal 1.3853 0.5419 2.556 0.0109 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.127 on 435 degrees of freedom  
## Multiple R-squared: 0.0148, Adjusted R-squared: 0.01253   
## F-statistic: 6.534 on 1 and 435 DF, p-value: 0.01092

summary(lm(jobrating ~ Factor1, data=kevin ))

##   
## Call:  
## lm(formula = jobrating ~ Factor1, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.2633 -3.1814 -0.1293 3.0510 15.7477   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.4005 0.2342 313.347 < 2e-16 \*\*\*  
## Factor1 1.7262 0.2467 6.997 9.91e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.897 on 435 degrees of freedom  
## Multiple R-squared: 0.1012, Adjusted R-squared: 0.09911   
## F-statistic: 48.96 on 1 and 435 DF, p-value: 9.91e-12

summary(lm(jobrating ~ Factor2, data=kevin ))

##   
## Call:  
## lm(formula = jobrating ~ Factor2, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.3504 -3.4562 -0.2528 3.5994 14.5224   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.40046 0.24704 297.114 <2e-16 \*\*\*  
## Factor2 -0.09178 0.26436 -0.347 0.729   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.164 on 435 degrees of freedom  
## Multiple R-squared: 0.000277, Adjusted R-squared: -0.002021   
## F-statistic: 0.1205 on 1 and 435 DF, p-value: 0.7286

summary(lm(jobrating ~ Factor3, data=kevin ))

##   
## Call:  
## lm(formula = jobrating ~ Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.6154 -3.3385 0.2012 3.4551 14.5465   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.4005 0.2462 298.104 <2e-16 \*\*\*  
## Factor3 0.4624 0.2659 1.739 0.0828 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.147 on 435 degrees of freedom  
## Multiple R-squared: 0.006903, Adjusted R-squared: 0.00462   
## F-statistic: 3.024 on 1 and 435 DF, p-value: 0.08276

chisq.test(table(kevin[,c("gend\_F","promote")]))

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(kevin[, c("gend\_F", "promote")])  
## X-squared = 10.976, df = 1, p-value = 0.0009232

chisq.test(table(kevin[,c("race","promote")]))

##   
## Pearson's Chi-squared test  
##   
## data: table(kevin[, c("race", "promote")])  
## X-squared = 6.7124, df = 2, p-value = 0.03487

summary(glm(promote ~ log\_sal, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ log\_sal, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1426 -0.9225 -0.8387 1.3937 1.7375   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.7524 2.6869 -2.513 0.0120 \*  
## log\_sal 0.5155 0.2280 2.261 0.0237 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 551.57 on 435 degrees of freedom  
## AIC: 555.57  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ Factor1, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2358 -0.9586 -0.7702 1.3227 2.1525   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.7299 0.1054 -6.924 4.38e-12 \*\*\*  
## Factor1 0.4988 0.1241 4.019 5.84e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 538.51 on 435 degrees of freedom  
## AIC: 542.51  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ Factor2, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ Factor2, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9208 -0.9057 -0.8943 1.4722 1.5004   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.68981 0.10143 -6.801 1.04e-11 \*\*\*  
## Factor2 0.02555 0.10858 0.235 0.814   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 556.72 on 435 degrees of freedom  
## AIC: 560.72  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ Factor3, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ Factor3, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9303 -0.9065 -0.8929 1.4685 1.5203   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.68994 0.10144 -6.801 1.04e-11 \*\*\*  
## Factor3 0.03921 0.10977 0.357 0.721   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 556.65 on 435 degrees of freedom  
## AIC: 560.65  
##   
## Number of Fisher Scoring iterations: 4

#controlling confounding for gender  
summary(lm(jobrating ~ gend\_F, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.7018 -3.0959 -0.0959 3.2982 15.2982   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.0959 0.3458 214.264 < 2e-16 \*\*\*  
## gend\_F -1.3941 0.4896 -2.847 0.00462 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.118 on 435 degrees of freedom  
## Multiple R-squared: 0.0183, Adjusted R-squared: 0.01604   
## F-statistic: 8.107 on 1 and 435 DF, p-value: 0.004619

summary(lm(jobrating ~ gend\_F \* log\_sal, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F \* log\_sal, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.5927 -3.2775 0.1679 3.3009 15.5192   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 52.8274 9.1905 5.748 1.71e-08 \*\*\*  
## gend\_F 0.9851 12.8061 0.077 0.939   
## log\_sal 1.8234 0.7874 2.316 0.021 \*   
## gend\_F:log\_sal -0.2256 1.0900 -0.207 0.836   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.072 on 433 degrees of freedom  
## Multiple R-squared: 0.04014, Adjusted R-squared: 0.03349   
## F-statistic: 6.036 on 3 and 433 DF, p-value: 0.0004927

summary(lm(jobrating ~ gend\_F + log\_sal, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F + log\_sal, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.6528 -3.2488 0.0982 3.3234 15.5341   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 54.2004 6.3534 8.531 2.45e-16 \*\*\*  
## gend\_F -1.6632 0.4923 -3.379 0.000794 \*\*\*  
## log\_sal 1.7057 0.5439 3.136 0.001829 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.066 on 434 degrees of freedom  
## Multiple R-squared: 0.04005, Adjusted R-squared: 0.03562   
## F-statistic: 9.053 on 2 and 434 DF, p-value: 0.0001406

summary(lm(jobrating ~ gend\_F \* Factor1, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F \* Factor1, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.4954 -3.0193 -0.0026 3.2399 15.8919   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.5832 0.3469 212.145 < 2e-16 \*\*\*  
## gend\_F -0.5443 0.4854 -1.121 0.263   
## Factor1 2.1381 0.4408 4.850 1.72e-06 \*\*\*  
## gend\_F:Factor1 -0.7385 0.5400 -1.368 0.172   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.889 on 433 degrees of freedom  
## Multiple R-squared: 0.1082, Adjusted R-squared: 0.1021   
## F-statistic: 17.52 on 3 and 433 DF, p-value: 9.461e-11

summary(lm(jobrating ~ gend\_F + Factor1, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F + Factor1, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.1071 -3.0263 -0.0113 3.2292 15.9964   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.7013 0.3363 219.164 < 2e-16 \*\*\*  
## gend\_F -0.6030 0.4840 -1.246 0.213   
## Factor1 1.6459 0.2548 6.459 2.84e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.894 on 434 degrees of freedom  
## Multiple R-squared: 0.1044, Adjusted R-squared: 0.1002   
## F-statistic: 25.29 on 2 and 434 DF, p-value: 4.085e-11

summary(lm(jobrating ~ gend\_F \* Factor2, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F \* Factor2, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.7729 -3.3259 0.1332 3.4569 15.3812   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.1368 0.3477 213.208 <2e-16 \*\*\*  
## gend\_F -1.4246 0.4923 -2.893 0.004 \*\*   
## Factor2 -0.4400 0.3703 -1.188 0.235   
## gend\_F:Factor2 0.5511 0.5269 1.046 0.296   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.121 on 433 degrees of freedom  
## Multiple R-squared: 0.02168, Adjusted R-squared: 0.01491   
## F-statistic: 3.199 on 3 and 433 DF, p-value: 0.0233

summary(lm(jobrating ~ gend\_F + Factor2, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F + Factor2, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.5946 -3.1836 0.0391 3.5420 15.1729   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.1115 0.3469 213.629 < 2e-16 \*\*\*  
## gend\_F -1.4253 0.4924 -2.895 0.00399 \*\*   
## Factor2 -0.1677 0.2635 -0.637 0.52464   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.121 on 434 degrees of freedom  
## Multiple R-squared: 0.01921, Adjusted R-squared: 0.01469   
## F-statistic: 4.251 on 2 and 434 DF, p-value: 0.01485

summary(lm(jobrating ~ gend\_F \* Factor3, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F \* Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.0769 -3.2551 0.0131 3.5465 15.1579   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.0856 0.3471 213.436 < 2e-16 \*\*\*  
## gend\_F -1.3204 0.4915 -2.687 0.00749 \*\*   
## Factor3 0.1092 0.3753 0.291 0.77113   
## gend\_F:Factor3 0.5613 0.5308 1.058 0.29085   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.11 on 433 degrees of freedom  
## Multiple R-squared: 0.02567, Adjusted R-squared: 0.01892   
## F-statistic: 3.802 on 3 and 433 DF, p-value: 0.01034

summary(lm(jobrating ~ gend\_F + Factor3, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F + Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.9199 -3.4535 0.1134 3.5937 15.2166   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.0592 0.3463 213.885 < 2e-16 \*\*\*  
## gend\_F -1.3206 0.4915 -2.687 0.00749 \*\*   
## Factor3 0.3898 0.2654 1.469 0.14265   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.111 on 434 degrees of freedom  
## Multiple R-squared: 0.02315, Adjusted R-squared: 0.01865   
## F-statistic: 5.143 on 2 and 434 DF, p-value: 0.006203

summary(lm(jobrating ~ gend\_F \* factor(race), data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F \* factor(race), data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.908 -3.099 -0.039 3.763 15.092   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.7476 0.5024 148.773 < 2e-16 \*\*\*  
## gend\_F -1.8393 0.7007 -2.625 0.00897 \*\*   
## factor(race)1 -0.7086 0.7682 -0.922 0.35681   
## factor(race)2 -2.2604 0.9587 -2.358 0.01883 \*   
## gend\_F:factor(race)1 -0.1011 1.0931 -0.092 0.92639   
## gend\_F:factor(race)2 2.5890 1.3572 1.908 0.05710 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.099 on 431 degrees of freedom  
## Multiple R-squared: 0.03435, Adjusted R-squared: 0.02314   
## F-statistic: 3.066 on 5 and 431 DF, p-value: 0.009878

summary(lm(jobrating ~ gend\_F + factor(race), data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ gend\_F + factor(race), data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.1135 -3.1135 -0.1135 3.4696 14.8865   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.5304 0.4319 172.547 < 2e-16 \*\*\*  
## gend\_F -1.4170 0.4893 -2.896 0.00397 \*\*   
## factor(race)1 -0.7426 0.5479 -1.355 0.17600   
## factor(race)2 -0.9740 0.6803 -1.432 0.15294   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.112 on 433 degrees of freedom  
## Multiple R-squared: 0.02488, Adjusted R-squared: 0.01812   
## F-statistic: 3.682 on 3 and 433 DF, p-value: 0.01217

summary(lm(jobrating ~ race, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.8019 -3.1081 0.1429 3.1981 14.1981   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.8019 0.3541 208.432 <2e-16 \*\*\*  
## race1 -0.6938 0.5522 -1.256 0.210   
## race2 -0.9447 0.6860 -1.377 0.169   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.156 on 434 degrees of freedom  
## Multiple R-squared: 0.005992, Adjusted R-squared: 0.001412   
## F-statistic: 1.308 on 2 and 434 DF, p-value: 0.2714

summary(lm(jobrating ~ race \* log\_sal, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race \* log\_sal, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.161 -3.484 0.089 3.250 14.238   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 67.0526 9.3970 7.136 4.11e-12 \*\*\*  
## race1 -16.3342 14.2800 -1.144 0.253   
## race2 -21.0309 17.3650 -1.211 0.227   
## log\_sal 0.5742 0.7989 0.719 0.473   
## race1:log\_sal 1.3341 1.2153 1.098 0.273   
## race2:log\_sal 1.7127 1.4780 1.159 0.247   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.125 on 431 degrees of freedom  
## Multiple R-squared: 0.02464, Adjusted R-squared: 0.01333   
## F-statistic: 2.178 on 5 and 431 DF, p-value: 0.05567

summary(lm(jobrating ~ race + log\_sal, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race + log\_sal, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.6573 -3.3744 0.1147 3.2686 14.2937   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.7398 6.3777 9.053 <2e-16 \*\*\*  
## race1 -0.6647 0.5490 -1.211 0.227   
## race2 -0.9179 0.6819 -1.346 0.179   
## log\_sal 1.3665 0.5418 2.522 0.012 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.124 on 433 degrees of freedom  
## Multiple R-squared: 0.02039, Adjusted R-squared: 0.0136   
## F-statistic: 3.004 on 3 and 433 DF, p-value: 0.03026

summary(lm(jobrating ~ race \* Factor1, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race \* Factor1, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.6422 -2.9557 -0.0133 3.1441 15.3324   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.8367 0.3364 219.485 < 2e-16 \*\*\*  
## race1 -0.6865 0.5247 -1.308 0.1914   
## race2 -1.1886 0.6564 -1.811 0.0709 .   
## Factor1 1.7577 0.3349 5.249 2.41e-07 \*\*\*  
## race1:Factor1 -0.1184 0.5347 -0.221 0.8248   
## race2:Factor1 0.2585 0.8342 0.310 0.7568   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.897 on 431 degrees of freedom  
## Multiple R-squared: 0.1093, Adjusted R-squared: 0.09896   
## F-statistic: 10.58 on 5 and 431 DF, p-value: 1.35e-09

summary(lm(jobrating ~ race + Factor1, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race + Factor1, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.6684 -2.9648 -0.0706 3.1822 15.3230   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.8364 0.3357 219.964 < 2e-16 \*\*\*  
## race1 -0.6836 0.5235 -1.306 0.1923   
## race2 -1.1600 0.6510 -1.782 0.0754 .   
## Factor1 1.7432 0.2465 7.071 6.2e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.887 on 433 degrees of freedom  
## Multiple R-squared: 0.1089, Adjusted R-squared: 0.1027   
## F-statistic: 17.64 on 3 and 433 DF, p-value: 8.071e-11

summary(lm(jobrating ~ race \* Factor2, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race \* Factor2, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.768 -3.419 0.156 3.267 14.135   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.80647 0.35596 207.345 <2e-16 \*\*\*  
## race1 -0.70107 0.55501 -1.263 0.207   
## race2 -0.98020 0.68985 -1.421 0.156   
## Factor2 -0.06997 0.37793 -0.185 0.853   
## race1:Factor2 0.02471 0.58192 0.042 0.966   
## race2:Factor2 -0.40412 0.78650 -0.514 0.608   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.17 on 431 degrees of freedom  
## Multiple R-squared: 0.007184, Adjusted R-squared: -0.004334   
## F-statistic: 0.6237 on 5 and 431 DF, p-value: 0.6818

summary(lm(jobrating ~ race + Factor2, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race + Factor2, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.7438 -3.2620 0.1253 3.3010 14.0887   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.8098 0.3548 208.016 <2e-16 \*\*\*  
## race1 -0.7089 0.5537 -1.280 0.201   
## race2 -0.9605 0.6875 -1.397 0.163   
## Factor2 -0.1209 0.2648 -0.457 0.648   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.16 on 433 degrees of freedom  
## Multiple R-squared: 0.006471, Adjusted R-squared: -0.0004129   
## F-statistic: 0.94 on 3 and 433 DF, p-value: 0.4212

summary(lm(jobrating ~ race \* Factor3, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race \* Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.937 -3.560 0.134 3.267 14.150   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.8219 0.3544 208.299 <2e-16 \*\*\*  
## race1 -0.7321 0.5517 -1.327 0.1852   
## race2 -1.1792 0.6952 -1.696 0.0906 .   
## Factor3 0.2479 0.3727 0.665 0.5063   
## race1:Factor3 0.3336 0.5814 0.574 0.5665   
## race2:Factor3 1.0809 0.8301 1.302 0.1936   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.142 on 431 degrees of freedom  
## Multiple R-squared: 0.01818, Adjusted R-squared: 0.00679   
## F-statistic: 1.596 on 5 and 431 DF, p-value: 0.1599

summary(lm(jobrating ~ race + Factor3, data=kevin))

##   
## Call:  
## lm(formula = jobrating ~ race + Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.0786 -3.3774 0.1031 3.3873 14.0990   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 73.8428 0.3537 208.787 <2e-16 \*\*\*  
## race1 -0.7507 0.5514 -1.361 0.174   
## race2 -1.0675 0.6870 -1.554 0.121   
## Factor3 0.5072 0.2668 1.901 0.058 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.14 on 433 degrees of freedom  
## Multiple R-squared: 0.01422, Adjusted R-squared: 0.007389   
## F-statistic: 2.082 on 3 and 433 DF, p-value: 0.1019

summary(glm(promote ~ gend\_F, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0288 -1.0288 -0.7706 1.3336 1.6487   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.3600 0.1373 -2.621 0.008762 \*\*   
## gend\_F -0.7022 0.2071 -3.391 0.000697 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 545.04 on 435 degrees of freedom  
## AIC: 549.04  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F \* log\_sal, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F \* log\_sal, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3389 -0.8933 -0.7425 1.2662 1.8314   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -10.5144 3.8701 -2.717 0.00659 \*\*  
## gend\_F 3.4270 5.6452 0.607 0.54381   
## log\_sal 0.8695 0.3308 2.629 0.00857 \*\*  
## gend\_F:log\_sal -0.3609 0.4786 -0.754 0.45078   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 535.60 on 433 degrees of freedom  
## AIC: 543.6  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F + log\_sal, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F + log\_sal, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2769 -0.9233 -0.7401 1.2785 1.9021   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.5300 2.7915 -3.056 0.002246 \*\*   
## gend\_F -0.8288 0.2148 -3.859 0.000114 \*\*\*  
## log\_sal 0.6997 0.2385 2.933 0.003355 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 536.17 on 434 degrees of freedom  
## AIC: 542.17  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F \* Factor1, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F \* Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3893 -0.8917 -0.7345 1.2352 2.0701   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.5187 0.1536 -3.376 0.000735 \*\*\*  
## gend\_F -0.4936 0.2196 -2.248 0.024593 \*   
## Factor1 0.5782 0.2014 2.870 0.004099 \*\*   
## gend\_F:Factor1 -0.2400 0.2587 -0.928 0.353578   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 531.35 on 433 degrees of freedom  
## AIC: 539.35  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F + Factor1, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F + Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2997 -0.9257 -0.7347 1.2573 2.1924   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.4759 0.1440 -3.306 0.000946 \*\*\*  
## gend\_F -0.5333 0.2136 -2.497 0.012537 \*   
## Factor1 0.4370 0.1277 3.423 0.000620 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 532.22 on 434 degrees of freedom  
## AIC: 538.22  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F \* Factor2, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F \* Factor2, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0497 -1.0155 -0.7652 1.3291 1.7007   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.36259 0.13809 -2.626 0.00865 \*\*   
## gend\_F -0.70629 0.20860 -3.386 0.00071 \*\*\*  
## Factor2 0.02728 0.14704 0.186 0.85281   
## gend\_F:Factor2 -0.08978 0.22287 -0.403 0.68707   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 544.86 on 433 degrees of freedom  
## AIC: 552.86  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F + Factor2, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F + Factor2, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0373 -1.0241 -0.7696 1.3328 1.6585   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.35892 0.13771 -2.606 0.009153 \*\*   
## gend\_F -0.70445 0.20816 -3.384 0.000714 \*\*\*  
## Factor2 -0.01179 0.11048 -0.107 0.915022   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 545.02 on 434 degrees of freedom  
## AIC: 551.02  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F \* Factor3, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F \* Factor3, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0523 -1.0213 -0.7696 1.3321 1.6741   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.35827 0.13802 -2.596 0.009437 \*\*   
## gend\_F -0.70164 0.20804 -3.373 0.000745 \*\*\*  
## Factor3 -0.01867 0.14920 -0.125 0.900444   
## gend\_F:Factor3 0.04489 0.22499 0.200 0.841860   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 545.00 on 433 degrees of freedom  
## AIC: 553  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F + Factor3, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F + Factor3, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0297 -1.0284 -0.7705 1.3335 1.6498   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.360106 0.137745 -2.614 0.008941 \*\*   
## gend\_F -0.702035 0.208169 -3.372 0.000745 \*\*\*  
## Factor3 0.001098 0.111626 0.010 0.992155   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 545.04 on 434 degrees of freedom  
## AIC: 551.04  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F \* race, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F \* race, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0717 -0.8576 -0.8135 1.2869 2.1219   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.25378 0.19865 -1.277 0.2014   
## gend\_F -0.62429 0.28931 -2.158 0.0309 \*  
## race1 -0.03390 0.30413 -0.111 0.9112   
## race2 -0.55715 0.39979 -1.394 0.1634   
## gend\_F:race1 -0.02412 0.45425 -0.053 0.9576   
## gend\_F:race2 -0.70485 0.69531 -1.014 0.3107   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 536.42 on 431 degrees of freedom  
## AIC: 548.42  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ gend\_F + race, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ gend\_F + race, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0896 -0.8163 -0.7998 1.2678 1.9500   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.20998 0.17452 -1.203 0.228895   
## gend\_F -0.71783 0.20906 -3.434 0.000596 \*\*\*  
## race1 -0.04787 0.22656 -0.211 0.832667   
## race2 -0.81166 0.31903 -2.544 0.010954 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 537.56 on 433 degrees of freedom  
## AIC: 545.56  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9501 -0.9501 -0.9416 1.4232 1.7727   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.56147 0.14281 -3.932 8.44e-05 \*\*\*  
## race1 -0.02212 0.22313 -0.099 0.9210   
## race2 -0.77682 0.31510 -2.465 0.0137 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 549.61 on 434 degrees of freedom  
## AIC: 555.61  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race \* log\_sal, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race \* log\_sal, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3035 -0.9382 -0.7510 1.3514 1.8016   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -5.5357 3.8684 -1.431 0.152  
## race1 -5.7340 5.9963 -0.956 0.339  
## race2 5.6024 7.9975 0.701 0.484  
## log\_sal 0.4228 0.3283 1.288 0.198  
## race1:log\_sal 0.4860 0.5087 0.955 0.339  
## race2:log\_sal -0.5426 0.6811 -0.797 0.426  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 542.17 on 431 degrees of freedom  
## AIC: 554.17  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race + log\_sal, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race + log\_sal, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1449 -0.9371 -0.8005 1.3441 2.0269   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.6837 2.7178 -2.459 0.0139 \*  
## race1 -0.0118 0.2246 -0.053 0.9581   
## race2 -0.7767 0.3167 -2.452 0.0142 \*  
## log\_sal 0.5203 0.2304 2.258 0.0239 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 544.42 on 433 degrees of freedom  
## AIC: 552.42  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race \* Factor1, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race \* Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3833 -0.9500 -0.7101 1.2819 2.0813   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.58088 0.14651 -3.965 7.35e-05 \*\*\*  
## race1 -0.04703 0.23275 -0.202 0.83988   
## race2 -0.86944 0.34273 -2.537 0.01119 \*   
## Factor1 0.42188 0.16003 2.636 0.00838 \*\*   
## race1:Factor1 0.22237 0.26896 0.827 0.40835   
## race2:Factor1 0.16285 0.47807 0.341 0.73338   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 529.56 on 431 degrees of freedom  
## AIC: 541.56  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race + Factor1, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race + Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3052 -0.9471 -0.7126 1.2600 2.1105   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.59255 0.14747 -4.018 5.87e-05 \*\*\*  
## race1 -0.01624 0.22850 -0.071 0.94336   
## race2 -0.83838 0.31934 -2.625 0.00866 \*\*   
## Factor1 0.51138 0.12385 4.129 3.64e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 530.28 on 433 degrees of freedom  
## AIC: 538.28  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race \* Factor2, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race \* Factor2, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0283 -0.9468 -0.8880 1.3985 1.9062   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.57083 0.14373 -3.972 7.14e-05 \*\*\*  
## race1 -0.01728 0.22421 -0.077 0.9386   
## race2 -0.78582 0.31940 -2.460 0.0139 \*   
## Factor2 0.11735 0.15281 0.768 0.4425   
## race1:Factor2 -0.18335 0.23510 -0.780 0.4355   
## race2:Factor2 -0.29552 0.36391 -0.812 0.4167   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 548.58 on 431 degrees of freedom  
## AIC: 560.58  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race + Factor2, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race + Factor2, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9606 -0.9477 -0.9336 1.4211 1.7833   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.56255 0.14301 -3.934 8.37e-05 \*\*\*  
## race1 -0.02011 0.22356 -0.090 0.928   
## race2 -0.77475 0.31541 -2.456 0.014 \*   
## Factor2 0.01601 0.10919 0.147 0.883   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 549.58 on 433 degrees of freedom  
## AIC: 557.58  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race \* Factor3, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race \* Factor3, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0943 -0.9460 -0.8411 1.3654 1.9231   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.55139 0.14334 -3.847 0.00012 \*\*\*  
## race1 -0.03063 0.22395 -0.137 0.89121   
## race2 -0.89339 0.33849 -2.639 0.00831 \*\*   
## Factor3 0.17011 0.15336 1.109 0.26733   
## race1:Factor3 -0.35418 0.23810 -1.488 0.13686   
## race2:Factor3 0.27167 0.39591 0.686 0.49259   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 545.82 on 431 degrees of freedom  
## AIC: 557.82  
##   
## Number of Fisher Scoring iterations: 4

summary(glm(promote ~ race + Factor3, data=kevin, family=binomial(link="logit")))

##   
## Call:  
## glm(formula = promote ~ race + Factor3, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9968 -0.9491 -0.9055 1.4066 1.7925   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.55685 0.14305 -3.893 9.91e-05 \*\*\*  
## race1 -0.02925 0.22357 -0.131 0.8959   
## race2 -0.79241 0.31638 -2.505 0.0123 \*   
## Factor3 0.06346 0.11050 0.574 0.5657   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 549.28 on 433 degrees of freedom  
## AIC: 557.28  
##   
## Number of Fisher Scoring iterations: 4

#model selection  
lm1 <- lm(jobrating~gend\_F + asian + other + log\_sal + Factor1 + Factor2 + Factor3, data=kevin) #baseline model  
summary(lm1)

##   
## Call:  
## lm(formula = jobrating ~ gend\_F + asian + other + log\_sal + Factor1 +   
## Factor2 + Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.1514 -2.9060 -0.0403 3.2054 15.1471   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 58.1257 5.9648 9.745 < 2e-16 \*\*\*  
## gend\_F -0.4794 0.4865 -0.985 0.32496   
## asian1 -0.8833 0.5089 -1.736 0.08331 .   
## other1 -1.5762 0.6363 -2.477 0.01363 \*   
## log\_sal 1.3702 0.5102 2.685 0.00752 \*\*   
## Factor1 2.0892 0.2672 7.818 4.20e-14 \*\*\*  
## Factor2 -0.5159 0.2509 -2.056 0.04039 \*   
## Factor3 1.3347 0.2724 4.900 1.36e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.727 on 429 degrees of freedom  
## Multiple R-squared: 0.1738, Adjusted R-squared: 0.1604   
## F-statistic: 12.9 on 7 and 429 DF, p-value: 4.479e-15

#backwise   
(backAIC <- step(lm1, direction = 'backward', data=kevin))

## Start: AIC=1365.58  
## jobrating ~ gend\_F + asian + other + log\_sal + Factor1 + Factor2 +   
## Factor3  
##   
## Df Sum of Sq RSS AIC  
## - gend\_F 1 21.70 9609.3 1364.6  
## <none> 9587.6 1365.6  
## - asian 1 67.34 9654.9 1366.6  
## - Factor2 1 94.47 9682.1 1367.9  
## - other 1 137.14 9724.7 1369.8  
## - log\_sal 1 161.17 9748.8 1370.9  
## - Factor3 1 536.63 10124.2 1387.4  
## - Factor1 1 1365.92 10953.5 1421.8  
##   
## Step: AIC=1364.57  
## jobrating ~ asian + other + log\_sal + Factor1 + Factor2 + Factor3  
##   
## Df Sum of Sq RSS AIC  
## <none> 9609.3 1364.6  
## - asian 1 65.71 9675.0 1365.5  
## - Factor2 1 90.78 9700.1 1366.7  
## - other 1 139.31 9748.6 1368.9  
## - log\_sal 1 144.87 9754.2 1369.1  
## - Factor3 1 594.61 10203.9 1388.8  
## - Factor1 1 1627.35 11236.6 1430.9

##   
## Call:  
## lm(formula = jobrating ~ asian + other + log\_sal + Factor1 +   
## Factor2 + Factor3, data = kevin)  
##   
## Coefficients:  
## (Intercept) asian1 other1 log\_sal Factor1   
## 58.9900 -0.8724 -1.5883 1.2761 2.1701   
## Factor2 Factor3   
## -0.5053 1.3825

(backBIC <- step(lm1, direction = 'backward', data=kevin, k=log(nrow(kevin))))

## Start: AIC=1398.22  
## jobrating ~ gend\_F + asian + other + log\_sal + Factor1 + Factor2 +   
## Factor3  
##   
## Df Sum of Sq RSS AIC  
## - gend\_F 1 21.70 9609.3 1393.1  
## - asian 1 67.34 9654.9 1395.2  
## - Factor2 1 94.47 9682.1 1396.4  
## <none> 9587.6 1398.2  
## - other 1 137.14 9724.7 1398.3  
## - log\_sal 1 161.17 9748.8 1399.4  
## - Factor3 1 536.63 10124.2 1415.9  
## - Factor1 1 1365.92 10953.5 1450.3  
##   
## Step: AIC=1393.13  
## jobrating ~ asian + other + log\_sal + Factor1 + Factor2 + Factor3  
##   
## Df Sum of Sq RSS AIC  
## - asian 1 65.71 9675.0 1390.0  
## - Factor2 1 90.78 9700.1 1391.2  
## <none> 9609.3 1393.1  
## - other 1 139.31 9748.6 1393.3  
## - log\_sal 1 144.87 9754.2 1393.6  
## - Factor3 1 594.61 10203.9 1413.3  
## - Factor1 1 1627.35 11236.6 1455.4  
##   
## Step: AIC=1390.03  
## jobrating ~ other + log\_sal + Factor1 + Factor2 + Factor3  
##   
## Df Sum of Sq RSS AIC  
## - Factor2 1 79.82 9754.8 1387.5  
## - other 1 92.68 9767.7 1388.1  
## <none> 9675.0 1390.0  
## - log\_sal 1 148.79 9823.8 1390.6  
## - Factor3 1 569.29 10244.3 1408.9  
## - Factor1 1 1611.53 11286.5 1451.3  
##   
## Step: AIC=1387.54  
## jobrating ~ other + log\_sal + Factor1 + Factor3  
##   
## Df Sum of Sq RSS AIC  
## - other 1 82.76 9837.6 1385.2  
## <none> 9754.8 1387.5  
## - log\_sal 1 152.99 9907.8 1388.3  
## - Factor3 1 499.96 10254.8 1403.3  
## - Factor1 1 1546.81 11301.6 1445.8  
##   
## Step: AIC=1385.15  
## jobrating ~ log\_sal + Factor1 + Factor3  
##   
## Df Sum of Sq RSS AIC  
## <none> 9837.6 1385.2  
## - log\_sal 1 154.50 9992.1 1385.9  
## - Factor3 1 464.25 10301.8 1399.2  
## - Factor1 1 1499.32 11336.9 1441.1

##   
## Call:  
## lm(formula = jobrating ~ log\_sal + Factor1 + Factor3, data = kevin)  
##   
## Coefficients:  
## (Intercept) log\_sal Factor1 Factor3   
## 57.931 1.317 2.059 1.176

glm1 <- glm(promote ~ gend\_F + asian + other + log\_sal + Factor1 + Factor2 + Factor3 + jobrating,  
 data=kevin, family=binomial(link="logit")) #baseline glm  
(glmbackAIC <- step(glm1, direction = 'backward', data=kevin))

## Start: AIC=527.87  
## promote ~ gend\_F + asian + other + log\_sal + Factor1 + Factor2 +   
## Factor3 + jobrating  
##   
## Df Deviance AIC  
## - asian 1 509.88 525.88  
## - Factor2 1 510.24 526.24  
## <none> 509.87 527.87  
## - Factor3 1 512.10 528.10  
## - jobrating 1 513.12 529.12  
## - log\_sal 1 515.77 531.77  
## - gend\_F 1 516.12 532.12  
## - other 1 517.41 533.41  
## - Factor1 1 519.27 535.27  
##   
## Step: AIC=525.88  
## promote ~ gend\_F + other + log\_sal + Factor1 + Factor2 + Factor3 +   
## jobrating  
##   
## Df Deviance AIC  
## - Factor2 1 510.24 524.24  
## <none> 509.88 525.88  
## - Factor3 1 512.10 526.10  
## - jobrating 1 513.19 527.19  
## - log\_sal 1 515.78 529.78  
## - gend\_F 1 516.12 530.12  
## - other 1 518.16 532.16  
## - Factor1 1 519.27 533.27  
##   
## Step: AIC=524.24  
## promote ~ gend\_F + other + log\_sal + Factor1 + Factor3 + jobrating  
##   
## Df Deviance AIC  
## - Factor3 1 512.16 524.16  
## <none> 510.24 524.24  
## - jobrating 1 513.81 525.81  
## - log\_sal 1 516.11 528.11  
## - gend\_F 1 516.36 528.36  
## - other 1 518.30 530.30  
## - Factor1 1 519.28 531.28  
##   
## Step: AIC=524.16  
## promote ~ gend\_F + other + log\_sal + Factor1 + jobrating  
##   
## Df Deviance AIC  
## <none> 512.16 524.16  
## - jobrating 1 517.03 527.03  
## - log\_sal 1 517.74 527.74  
## - Factor1 1 519.31 529.31  
## - other 1 519.40 529.40  
## - gend\_F 1 519.76 529.76

##   
## Call: glm(formula = promote ~ gend\_F + other + log\_sal + Factor1 +   
## jobrating, family = binomial(link = "logit"), data = kevin)  
##   
## Coefficients:  
## (Intercept) gend\_F other1 log\_sal Factor1   
## -10.74241 -0.61656 -0.79663 0.57862 0.34621   
## jobrating   
## 0.04929   
##   
## Degrees of Freedom: 436 Total (i.e. Null); 431 Residual  
## Null Deviance: 556.8   
## Residual Deviance: 512.2 AIC: 524.2

(glmbackBIC <- step(glm1, direction = 'backward', data=kevin, k=log(nrow(kevin))))

## Start: AIC=564.59  
## promote ~ gend\_F + asian + other + log\_sal + Factor1 + Factor2 +   
## Factor3 + jobrating  
##   
## Df Deviance AIC  
## - asian 1 509.88 558.52  
## - Factor2 1 510.24 558.88  
## - Factor3 1 512.10 560.74  
## - jobrating 1 513.12 561.76  
## - log\_sal 1 515.77 564.41  
## <none> 509.87 564.59  
## - gend\_F 1 516.12 564.75  
## - other 1 517.41 566.05  
## - Factor1 1 519.27 567.91  
##   
## Step: AIC=558.52  
## promote ~ gend\_F + other + log\_sal + Factor1 + Factor2 + Factor3 +   
## jobrating  
##   
## Df Deviance AIC  
## - Factor2 1 510.24 552.80  
## - Factor3 1 512.10 554.66  
## - jobrating 1 513.19 555.75  
## - log\_sal 1 515.78 558.34  
## <none> 509.88 558.52  
## - gend\_F 1 516.12 558.68  
## - other 1 518.16 560.72  
## - Factor1 1 519.27 561.83  
##   
## Step: AIC=552.8  
## promote ~ gend\_F + other + log\_sal + Factor1 + Factor3 + jobrating  
##   
## Df Deviance AIC  
## - Factor3 1 512.16 548.64  
## - jobrating 1 513.81 550.29  
## - log\_sal 1 516.11 552.59  
## <none> 510.24 552.80  
## - gend\_F 1 516.36 552.84  
## - other 1 518.30 554.78  
## - Factor1 1 519.28 555.75  
##   
## Step: AIC=548.64  
## promote ~ gend\_F + other + log\_sal + Factor1 + jobrating  
##   
## Df Deviance AIC  
## - jobrating 1 517.03 547.43  
## - log\_sal 1 517.74 548.14  
## <none> 512.16 548.64  
## - Factor1 1 519.31 549.71  
## - other 1 519.40 549.80  
## - gend\_F 1 519.76 550.16  
##   
## Step: AIC=547.43  
## promote ~ gend\_F + other + log\_sal + Factor1  
##   
## Df Deviance AIC  
## <none> 517.03 547.43  
## - log\_sal 1 524.09 548.41  
## - other 1 525.13 549.45  
## - gend\_F 1 525.79 550.11  
## - Factor1 1 528.85 553.17

##   
## Call: glm(formula = promote ~ gend\_F + other + log\_sal + Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Coefficients:  
## (Intercept) gend\_F other1 log\_sal Factor1   
## -7.8138 -0.6550 -0.8359 0.6402 0.4248   
##   
## Degrees of Freedom: 436 Total (i.e. Null); 432 Residual  
## Null Deviance: 556.8   
## Residual Deviance: 517 AIC: 527

summary(backAIC)

##   
## Call:  
## lm(formula = jobrating ~ asian + other + log\_sal + Factor1 +   
## Factor2 + Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.2112 -2.9969 -0.0409 3.0110 14.9601   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 58.9900 5.8998 9.999 < 2e-16 \*\*\*  
## asian1 -0.8724 0.5087 -1.715 0.0871 .   
## other1 -1.5883 0.6361 -2.497 0.0129 \*   
## log\_sal 1.2761 0.5012 2.546 0.0112 \*   
## Factor1 2.1701 0.2543 8.534 2.46e-16 \*\*\*  
## Factor2 -0.5053 0.2507 -2.015 0.0445 \*   
## Factor3 1.3825 0.2680 5.158 3.81e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.727 on 430 degrees of freedom  
## Multiple R-squared: 0.172, Adjusted R-squared: 0.1604   
## F-statistic: 14.88 on 6 and 430 DF, p-value: 1.741e-15

summary(backBIC)

##   
## Call:  
## lm(formula = jobrating ~ log\_sal + Factor1 + Factor3, data = kevin)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.0441 -3.0968 0.1825 2.9725 15.9120   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.9308 5.9366 9.758 < 2e-16 \*\*\*  
## log\_sal 1.3173 0.5052 2.608 0.00943 \*\*   
## Factor1 2.0588 0.2534 8.124 4.77e-15 \*\*\*  
## Factor3 1.1758 0.2601 4.520 7.98e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.767 on 433 degrees of freedom  
## Multiple R-squared: 0.1523, Adjusted R-squared: 0.1464   
## F-statistic: 25.93 on 3 and 433 DF, p-value: 1.919e-15

anova( backBIC,backAIC, test = "F")

## Analysis of Variance Table  
##   
## Model 1: jobrating ~ log\_sal + Factor1 + Factor3  
## Model 2: jobrating ~ asian + other + log\_sal + Factor1 + Factor2 + Factor3  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 433 9837.6   
## 2 430 9609.3 3 228.29 3.4053 0.01767 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(glmbackAIC)

##   
## Call:  
## glm(formula = promote ~ gend\_F + other + log\_sal + Factor1 +   
## jobrating, family = binomial(link = "logit"), data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4975 -0.9050 -0.6408 1.1882 2.1217   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -10.74241 3.18948 -3.368 0.000757 \*\*\*  
## gend\_F -0.61656 0.22529 -2.737 0.006204 \*\*   
## other1 -0.79663 0.31010 -2.569 0.010201 \*   
## log\_sal 0.57862 0.24722 2.341 0.019257 \*   
## Factor1 0.34621 0.13331 2.597 0.009405 \*\*   
## jobrating 0.04929 0.02251 2.190 0.028540 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 512.16 on 431 degrees of freedom  
## AIC: 524.16  
##   
## Number of Fisher Scoring iterations: 4

summary(glmbackBIC)

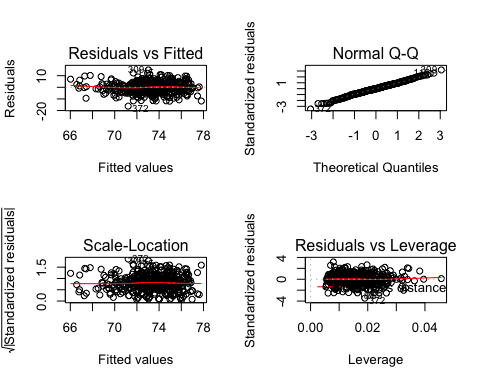
##   
## Call:  
## glm(formula = promote ~ gend\_F + other + log\_sal + Factor1, family = binomial(link = "logit"),   
## data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4115 -0.9173 -0.6721 1.2004 2.1833   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -7.8138 2.8538 -2.738 0.006181 \*\*   
## gend\_F -0.6550 0.2233 -2.933 0.003358 \*\*   
## other1 -0.8359 0.3086 -2.708 0.006765 \*\*   
## log\_sal 0.6402 0.2440 2.623 0.008706 \*\*   
## Factor1 0.4248 0.1288 3.298 0.000975 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 517.03 on 432 degrees of freedom  
## AIC: 527.03  
##   
## Number of Fisher Scoring iterations: 4

anova(glmbackBIC,glmbackAIC, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: promote ~ gend\_F + other + log\_sal + Factor1  
## Model 2: promote ~ gend\_F + other + log\_sal + Factor1 + jobrating  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 432 517.03   
## 2 431 512.16 1 4.8748 0.02725 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#best model backAIC, glmbackAIC

#model assessment   
#for lm  
par(mfrow = c(2,2))  
plot(backAIC)



#for glm  
#chisq test  
anova(glmbackAIC, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: promote  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 436 556.77   
## gend\_F 1 11.7390 435 545.04 0.0006120 \*\*\*  
## other 1 7.4353 434 537.60 0.0063958 \*\*   
## log\_sal 1 8.7523 433 528.85 0.0030921 \*\*   
## Factor1 1 11.8146 432 517.03 0.0005877 \*\*\*  
## jobrating 1 4.8748 431 512.16 0.0272509 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

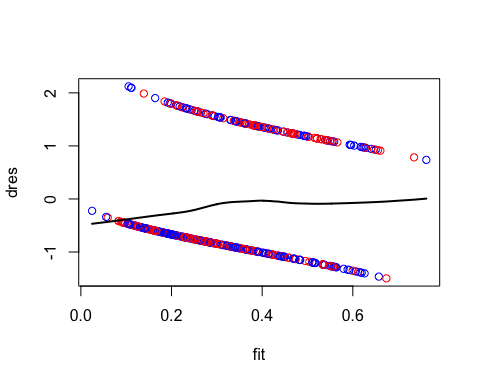
#HL test  
library(MKmisc)

## Warning: package 'MKmisc' was built under R version 3.4.4

HLgof.test(fit = fitted(glmbackAIC), obs = kevin$promote)

## $C  
##   
## Hosmer-Lemeshow C statistic  
##   
## data: fitted(glmbackAIC) and kevin$promote  
## X-squared = 9.696, df = 8, p-value = 0.287  
##   
##   
## $H  
##   
## Hosmer-Lemeshow H statistic  
##   
## data: fitted(glmbackAIC) and kevin$promote  
## X-squared = 7.9687, df = 8, p-value = 0.4365

dres <- residuals(glmbackAIC, type = "deviance")  
fit <- fitted(glmbackAIC)  
  
plot(fit, dres, col = c("red", "blue"))  
lines(lowess(fit,dres),col="black",lwd=2)



pres <- residuals(glmbackAIC, type = "pearson")  
(disp <- sum((pres^2)/431))

## [1] 0.9939851

summary(glmbackAIC, dispersion = disp)

##   
## Call:  
## glm(formula = promote ~ gend\_F + other + log\_sal + Factor1 +   
## jobrating, family = binomial(link = "logit"), data = kevin)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4975 -0.9050 -0.6408 1.1882 2.1217   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -10.74241 3.17987 -3.378 0.000729 \*\*\*  
## gend\_F -0.61656 0.22461 -2.745 0.006050 \*\*   
## other1 -0.79663 0.30916 -2.577 0.009975 \*\*   
## log\_sal 0.57862 0.24648 2.348 0.018896 \*   
## Factor1 0.34621 0.13291 2.605 0.009192 \*\*   
## jobrating 0.04929 0.02244 2.196 0.028064 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 0.9939851)  
##   
## Null deviance: 556.77 on 436 degrees of freedom  
## Residual deviance: 512.16 on 431 degrees of freedom  
## AIC: 524.16  
##   
## Number of Fisher Scoring iterations: 4