The Interactive Effect of Various Socio-economic Factors and the Participation of Job Training Program on Mental Health and Reemployment Status in Unemployed Individuals

Dongyang Cai, Yilin Song, and Zhaoliang Zhou

05/13/2018

### 1. Introduction

In this project, we will explore the effect of both socio-economic factors and participation in a job training program on reemployment and one’s mental health, specifically depression. Even though there is an increasing number of job training programs, social and economic stresses are also important factors, which could potentially influence reemployment rate. This thus prompts us to investigate these underlying phenomena. In order to accomplish this task, we examined the data that was collected as a part of the Job Search Intervention Program evaluation, which conducted at the University of Michigan in 1992. We focused on the individual’s status of participation in the program along with socio-economic explanatory variables such as gender, age, marital status, occupation, education, and job-search efficacy in order to explain the success of reemployment and the improvement of depression score, the latter of which was measured quantitatively.

Respondents were recruited from four offices of the Michigan Employment Security Commission (MESC) in southeastern Michigan in 1992 to participate in the Job Search Training Program. The data was originally collected from 1,801 non-randomly sampled unemployed adult respondents who participated in the program. The dataset we are using is a condensed version of the original dataset and contains 899 random observation entries of 17 different variables. The original study only focused on the efficacy of a job training intervention on unemployed workers and conducted different tests such as structural equation modeling and two-way analysis of covariance. We deleted certain variables which we are not interested in from the dataset and obtained a new dataset with 899 entries and 10 variables.

For our response variables, we decided to use the binary variable- “work1”- which indicates whether the person had been employed and a quantitative variable- “depress”- which shows the improvement of depression score. The variable “work1” has 2 levels: unemployed and employed. The depression variables are continuous quantitative with values from 1-5. This information was gathered by asking each participant 11 questions about their own experiences with depression, for which they answered on a categorical scale from 1-5 (1 meaning “not at all” and 5 “extremely”), and the answers to the 11 questions were averaged. This method is based on the John Hopkins Symptoms Checklist, which had 25 questions about overall mental health, with answers ranging from 1-4, which represent “not at all,” “a little,” “quite a bit,” and “extremely,” respectively. The depression score was calculated by averaging the answers for 15 questions specifically about depression. For our research, we will focus on the improvement of depressive symptoms between pretreatment (depress1) and post-treatment (depress2) rather than using each of them as single predictors. To show the improvement of depressive symptoms we subtract pretreatment scores from post-treatment scores and get our new response variable called “depress”. For this variable, a positive value indicates the depressive symptoms have become worse after the program, and a negative value indicates the depressive symptoms have become better after this job training program. In other words, if a positive value is found, the depressive symptoms increased, and treatment was not effective.

For explanatory variables, we only focused on the socio-economic factors of our interests including “occupation”, “age”, “marital status”, “sex”, “education”, “comply”, and “job\_seek”. A survey was used to gather demographic information of age, sex, education, marital status, and occupation. The “age” variable is continuous quantitative. The “sex” variable is categorical, with the 2 categories being male and female. The “education” variable has 5 categories: less than high school, high school, community college, bachelors, and graduate. The “marital status” variable is categorical, with 5 categories: married, never married, separated, divorced, and widowed. The “occupation” variable is categorical with 7 categories: professionals, operatives, managerial, sales workers, craftsmen, clerical, and laborers/service. The variable “comply” indicates whether participants actually participated in the JOBS II program (1=participation, 0=no participation). The variable “job\_seek” measures job search efficacy using a six-item index on a 5-point scale that shows the degree of confidence in being able to successfully perform the essential job-search activities. There are no missing values in our dataset, and we did not manipulate any data.

Past findings about the effect of job training program suggest that depression is significantly improved by using the reemployment status as a secondary outcome in the study (Mattei, 2013). Also, as suggested in a prior study, “workers receiving skill training are likely to have higher reemployment probability than those who have received only job search assistance or who did not participate in any job training program” (Ting, 1991). Furthermore, it makes intuitive sense that worsening socio-economic environment increases depression. More specifically, an education a survey performed in Europe suggests that “for each additional year of education, the odds of being depressed decreased by 3%” (Aislinne, 2016). Regarding family situations, family separation can have a significantly negative impact on depression status (Lorant et al., 2007). Research has also shown that older male workers have higher reemployment probability (Ting, 1991).

Based on past findings, we hypothesized that socio-economic factors and the participation of job training program have significant influences on one’s mental health and reemployment status.

### 2. Results

#### “depress” as response variable

According to the statistics summary and histogram (Appendix A1) for “depress”, we found that the mean of the improvement of depression scores overall in this sample is -0.1286 with a 0.653 standard deviation, and the distribution is approximately bell-curved and follows a normal distribution.

According to our models with single predictors, only “comply” (Appendix B6) and “job\_seek” (Appendix B13) have significant relationships with “depress”.

We were first interested in the interactive impact of whether participating in the program and seeking for jobs have impacts on depression status, so we created a model with “comply + job\_seek + comply:job\_seek” as predictors. Based on the summary output (Appendix C4), the overall model has a p-value of 0.001, which is significant on a 0.05 significance level. However, when looking at the p-values for the coefficients of individual predictors, both the interaction term (p-value 0.31) and “comply” (p-value 0.485) are not significant. Thus, we decided to run a backward selection to test for the interaction term and “comply” (Appendix C6). The results of backward selection suggest us to omit the interaction term. Therefore, we reach a conclusion that there is no interactive impact of participation in job training program and seeking jobs on depression status.

When only including predictors “comply” and “job\_seek”, the overall model has a p-value 0.0001, which is significant on a 0.05 level. However, the individual variable “comply” still has an insignificant p-value 0.079, so we decided to run a Nested-F test to see whether including the term “comply” improves our model. The Nested-F test summary output (Appendix C3)indicates the term “comply” is not necessary for our model.

The new model with only “comply” and “job\_seek” has an overall p-value of nearly 0, which is considered statistically significant (Appendix C2). When looking at the diagnostic plot (Appendix C1) for the linear regression model between “depress” and “job\_seek”, the conditions for linearity, normality, and equal variance are all met.

The predictor variable “job\_seek” has a coefficient -0.1. Since the model has a statistically significant p-value on 0.05 level, we could say that the evidence is strong enough to support that on average for every additional unit increase of job seek self-efficacy, the post-treatment depression score would decrease by 0.1, indicating a decrease in depressive symptoms level from pre-treatment. Based on the 95% confidence interval output, the slope coefficient has a 95% CI (-0.16, -0.04). This indicates we are 95% confident that for every additional increase in job-seek self-efficacy, the depressive symptoms score would decrease from 0.04 to 0.16. However, considering the scale of “job\_seek” variable which is from 1-5, the largest impact it has on depression score is 0.5. Therefore, even though job seek-efficacy is a significant predictor of the improvement or depression scores, the change is not substantial in practical.

#### work as response variable

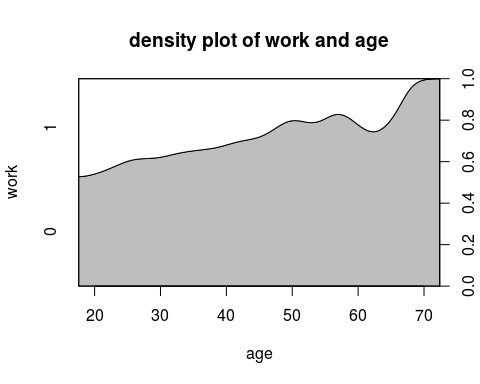
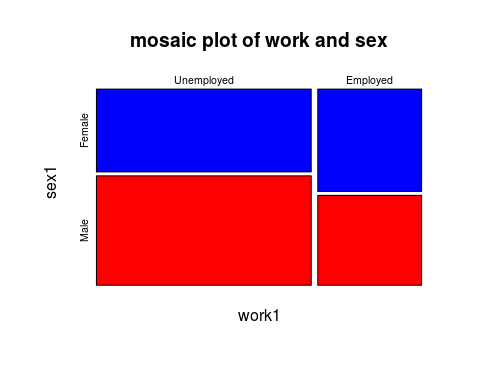
According to the table (Appendix A7) of our second response variable “work1”, we found there were 293 participants reemployed (32.6%), and the rest 606 did not (67.4%).

According to our summaries with single predictors (Appendix B10 & B12), only “age” and “gender” have significant relationships with “work1”.

We first created a conditioned density plot of “age” and “work1” (Appendix D4), we found that the probability of being reemployed decreases as age increases. Also, the mean age of reemployed group (34.21) is 4.5 years younger than the mean age of unemployed group (37.85). From the logistic regression model (Appendix B24) between work and age, the p-value is around 0, which is smaller than 0.05, which suggest that there is a relationship between age and the probability of reemployment. Furthermore, according to the table of “gender”(Appendix A3) , the proportion of male being employed (17%) is more than the proportion of female being employed (15%). Since the p-value of the chi-squared test (Appendix B12) is 0.005, which is smaller than 0.05, we have statistically significant evidence to suggest that males are more likely to be re-employed than females. All the other explanatory variables including “occupation”, “marital”, “comply”, “education” and “job\_seek” do not have significant relationships with “work1”, since their individual p-values appeared to be all larger than 0.05 which is our significant level.

Therefore, in order to test the interactive impact of gender and age on reemployment status, we created a logistic regression model with “age” + gender + age:gender” as predictors (Appendix D2). Surprisingly, the overall p-value of our model is 0.116, which is insignificant, and both “sex” (p-value 0.759) and the interaction term “age:gender” (p-value 0.659) are insignificant as well. The predictor “age” has a p-value of 0.003, which is the only significant predictor. Since the p-value for the interaction term is the biggest, we decided to run a backward selection (Appendix D3) for the overall model. We found that without the interaction term, the model’s AIC becomes smaller, which suggests us to omit the interaction term and run a new model. Therefore, we reach a conclusion that there is no interactive impact of age and gender on the probability of reemployment.

The new model with only “age” and “gender” (Appendix D1) has an overall p-value of 0.009, which is considered statistically significant. According to the individual tests for the coefficients in the model, “age” and “gender” are significant predictors with p-values 0 and 0.006 respectively. The model equation is work=0.737-0.034(age)-0.400(gender) This provides us an idea that if a person is 10 years younger, the odds of being reemployed will be 5.814 times of the odds of another person but with the same gender. Also, controlling for age, if the participant is female, the odds of being reemployed for females is only 0.67 times the odds for males. Therefore, the result reflects our previous findings from EDA perfectly, and also in real life older female participants are harder to be employed.



| Model | Response | Coefficients | Estimate | P-value | Test statistics | CI | R-squared/ AIC |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model 2  (final model) | depress | intercept | 0.282 | 0.021 | 2.304 | [0.042, 0.521] | 0.012 |
| job\_seek | -0.101 | 0.001 | 11.64 | [-0.160, -0.040] |
| Equation: depress = 0.282 - 0.101 (job\_seek) | | | | | | |
| Model 8  (final model) | work | intercept | 0.738 | 0.009 | 2.604 | [1.204, 3.659] | 1109.7 |
| age | -0.034 | 0.000 | -4.676 | [0.953, 0.980] |
| sex | -0.401 | 0.006 | -2.768 | [0.504, 0.889] |
| Equation: work = 0.738 - 0.034 (age) - 0.401 (sex) | | | | | | |

### 3. Discussion

Overall, our results only demonstrate the effect of socio-economic factors on reemployment and on one’s mental health. Specifically, out of the socio-economic factors that we examined (job seek self-efficacy, education, marital status, gender, occupation, and age) only three factors (age, gender, job seek self-efficacy) were suggested to be significantly associated with depression and reemployment. The participation of the program has no impact on reemployment status, and the impact of the participation on the improvement of depression score became insignificant after adjusting for the job seek efficacy.

Our findings showed that age and gender significantly affect reemployment status. Older participants are less likely to be reemployed, and female participants also have a low probability to be reemployed. However, from one of the early studies in 1991, researchers have shown that older male workers have higher reemployment probability (Ting, 1991). Our results seem to have some contradiction with this study, which is possibly due to our relatively small sample size.

Although education and participation in job training program appeared to be insignificant predictors for reemployment status and depression in our data, some publications on this subject suggest otherwise. Ting suggested in 1991 that “workers receiving skill training are likely to have higher reemployment probability than those who have received only job search assistance or who did not participate in any job training program” (Ting, 1991). Another study in 2016 showed that “for each additional year of education, the odds of being depressed decreased by 3 %…” (Aislinne, 2016). Even though it makes intuitive sense and our EDA supports this trend, our study did not yield a statistically significant p-value to support this intuition.

Our findings also showed that there is a relationship between job search self-efficacy and depression scores. Job search self-efficacy is negatively related to depression scores, indicating that if one has more job search self-efficacy, this person would have lower depression scores from post-treatment compared to pre-treatment. Our findings yield similar results to one study conducted in 2011, which was exploring the relationships between self-efficacy and symptoms of depression (Tahmassian et al., 2011). Tahmassian et al. concluded that “different domains of self-efficacy and symptoms of affective disorders are significantly correlated”. They investigated emotional self-efficacy, physical self-efficacy, social self-efficacy, and total self-efficacy in their research, and all the predictors had significant negative relationships with depressive symptoms. Although their research focused on a different domain of self-efficacy and mostly focused on high school students, their results still have some convincing evidence on the negative relationship between self-efficacy and depression symptoms.   
  
Our results are not generalizable to the entire US population. Since respondents were recruited and were only unemployed individuals who participated in job search training program. Also, our data are not randomly sampled. However, since our dataset with 899 instances was a small random sample from the original data set, we can generalize our results to the original data set. Since “comply” is the only variable that was measured in an experimental setting, so we can only make causal conclusions for “comply” but not the other socio-economic factors. However, we found no relationships between “comply” with either depression symptoms or reemployment status. Therefore, there is no causal conclusions in our study.

There are several limitations in our data set. For example, the 0’s in the “comply” variable include both the people in the control group and the people in the treatment group who chose not to participate. We treat them equally when processing the model. However, they have potential differences because people who are assigned to treatment group then decided to drop out the training program should be different from those people who decided not participate initially. The results we found may be affected by the fact that some people in treatment group decided not to participate but not the efficiency of the job training program.

The depression score scale used in the study was not very clear at defining each number on the scale since we only know 1-5 as from “not at all” to “extremely”. This makes a number such as 2 not very meaningful since we could not accurately describe what a score of 2 means, and respondents may have different interpretations of the scales when answering. In addition, our depression score variable is not perfectly quantitative, because it is only quantitative after taking the average of values for an ordinal categorical variable (1,2,3,4,5). We used the improvement of depression scores instead of only the post-treatment scores as the response variable. However, pretreatment scores may still cause some biases in our model. For example, it would be hard for an extremely depressed person to change his depression status, but once there is a small improvement, he might refer that as a “big” change which would be different from what a slightly depressed person think of a “big” improvement. Also, there is not a lot of space of “not at all” depressed person to improve as an “extremely” depressed person has.

Furthermore, the R2 in our final model for “depress” and job seek-efficacy is only 1.17%, and the AIC in our final model between reemployment status and age and gender is 1109, which means our models barely explain our data. Depression scores and reemployment status could be explained by a lot of other variables, so it is expected our models with a single or a couple predictors would have a low R2. Even though there are a lot of variabilities that can’t be explained by a model, we still get the information that job seek-efficacy is related to depress, and age and gender are related to reemployment status.

Our sample size (899) is reasonable considering our research topic and the tests we ran. Our data come from a randomized field experiment, which is the strength of our study. Also, there is no missing values in our dataset. There are some possible confounding variables including health situations, employment history or geographic areas, which are not even available in our dataset and, thus, may cause some biases in our study.

Further research could be done with a larger dataset, which is ideally a random sample from the population so that we can make predictions towards the whole population. We expect clearer separation between the control group and the treatment group and eliminate situations like participants are assigned to treatment group but decide to drop out. We also expect to have a better definition of response variables like “depress” in our data set, and all separation within the response variables are expected to be clear and meaningful. More importantly, we expect to have more socio-economic variables to make sure we have at least a handful of significant relationships between explanatory and response variables. It would be beneficial to all of us if there are more studies to investigate the underlying factors that may contribute to one’s depression and reemployment probability.

### 4. Reference

Aislinne F., Stefanos T., Ai K., Somnath C., Matilde L., Jose Luis A., Beata T., Seppo K., Christine R., Josep Maria H.; “The role of socio-economic status in depression: results from the COURAGE (aging survey in Europe)”; BMC Public Health, BMC series – open, inclusive and trusted 2016 16:1098.

Alessandra Mattei., Fan Li., Fabrizia Mealli. “Exploiting Multiple Outcomes in Bayesian Principal Stratification Analysis with Application to the Evaluation of a Job Training Program.” Institute of Mathematical Statistics. Vol. 7, No. 4 (December 2013), pp. 2336-2360.

Marios Michaelides; “Are Reemployment Services Effective in Periods of High Unemployment? Experimental Evidence from the UI System”; \*Lecturer, Department of Economics, University of Cyprus; Research Fellow, IMPAQ International.

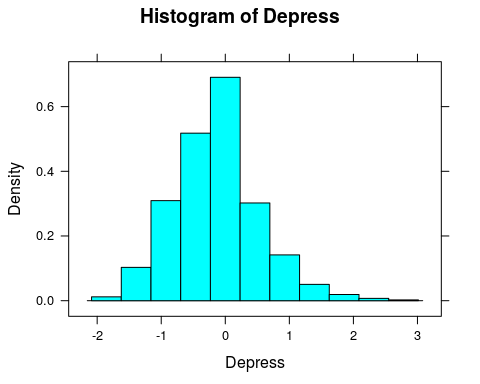
Tahmassian, Karineh., Moghadam, Niloufar J. “Relationship Between Self-Efficacy and Symptoms of Anxiety, Depression, Worry and Social Avoidance in a Normal Sample of Students”. Iranian Journal of Psychiatry and Behavioral Sciences. 2011 Autumn-Winter; 5(2): 91–98.

Ting, Yuan. “The Impact of Job Training Programs on the Reemployment Probability of Disclosed Workers.” Policies Study Review. Spring/Summer 1991, Vol.10, No. ⅔. Pp. 31-44.

### 5. Appendix: You can include residual plots, code, etc here.

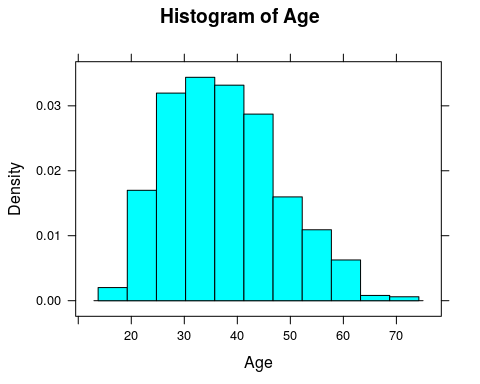
#### Appendix A: Single variables

##### A1. Depression



## mean standard deviation sample size missing  
## -0.129 0.653 899.000 0.000

##### A2. Age



## mean standard deviation sample size missing  
## -0.129 0.653 899.000 0.000

##### A3. Sex

##   
## 0 1   
## 0.4638487 0.5361513

##### A4. Marital Status

##   
## divrcd married nevmarr separtd widowed   
## 0.18131257 0.45383760 0.31034483 0.03337041 0.02113459

##### A5. Education

##   
## bach gradwk highsc lt-hs somcol   
## 0.16240267 0.12458287 0.30255840 0.05561735 0.35483871

##### A6. Comply

##   
## 0 1   
## 0.5862069 0.4137931

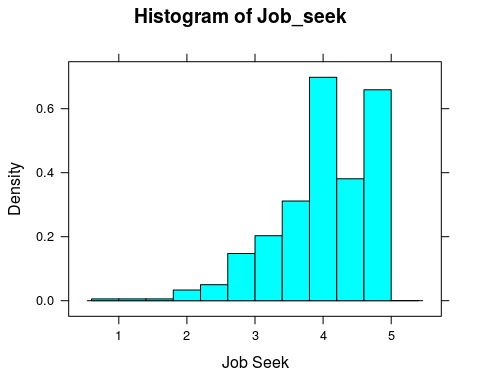
##### A7. Work1

##   
## psyemp psyump   
## 0.3259177 0.6740823

##### A8. Occupation

##   
## clerical/kindred craftsmen/foremen/kindred   
## 0.24137931 0.10789766   
## laborers/service wrks manegerial   
## 0.09343715 0.18687430   
## operatives/kindred wrks professionals   
## 0.10344828 0.19466073   
## sales workers   
## 0.07230256

##### A9. Job-seek



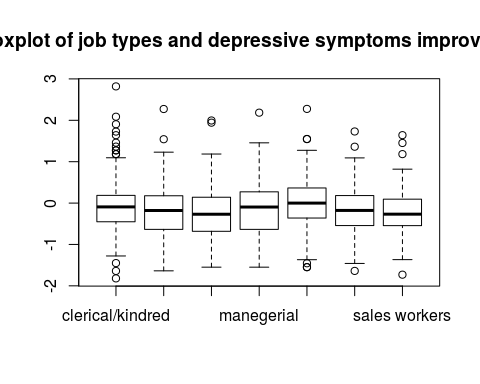
## mean standard deviation sample size missing  
## 4.043 0.728 899.000 0.000

#### Appendix B: Two variable EDA

##### B1. occupation and depress

## Mean sd n  
## clerical/kindred -0.070 0.660 217.000  
## craftsmen/foremen/kindred -2.040 0.640 97.000  
## laborers/service wrks -0.187 0.720 84.000  
## manegerial -0.123 0.640 168.000  
## operatives/kindred wrks 0.008 0.710 93.000  
## professionals -0.170 0.590 175.000  
## sales workers -0.200 0.610 65.000

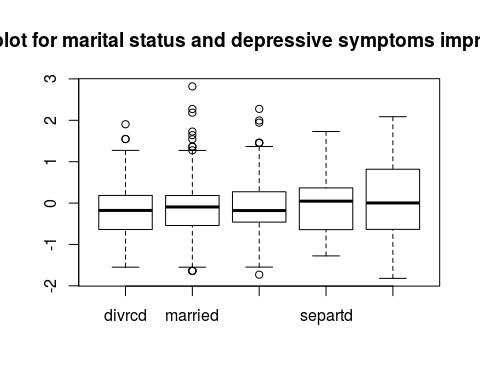
## Df Sum Sq Mean Sq F value Pr(>F)  
## occp 6 4 0.6671 1.57 0.153  
## Residuals 892 379 0.4249



##### B2. marital and depress

## Mean sd n  
## Divorced -1.82000 0.65700 163.00000  
## Married -0.12300 0.64500 408.00000  
## Never Married -0.13100 0.62700 279.00000  
## Separated -0.00018 0.71300 30.00000  
## Widowed 0.40600 0.99600 19.00000

## Df Sum Sq Mean Sq F value Pr(>F)  
## marital 4 1.5 0.3804 0.891 0.468  
## Residuals 894 381.5 0.4267



##### B3. occupation and work

## occp  
## work1 clerical/kindred craftsmen/foremen/kindred laborers/service wrks  
## psyemp 53 37 32  
## psyump 164 60 52  
## occp  
## work1 manegerial operatives/kindred wrks professionals sales workers  
## psyemp 55 37 57 22  
## psyump 113 56 118 43

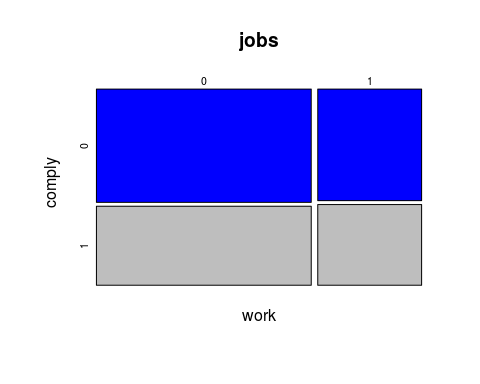
##   
## Pearson's Chi-squared test  
##   
## data: work.table  
## X-squared = 11.347, df = 6, p-value = 0.07822

##### B4. marital and work

## marital  
## work1 divrcd married nevmarr separtd widowed  
## psyemp 58 120 102 10 3  
## psyump 105 288 177 20 16

##   
## Pearson's Chi-squared test  
##   
## data: mar.table  
## X-squared = 6.9897, df = 4, p-value = 0.1364

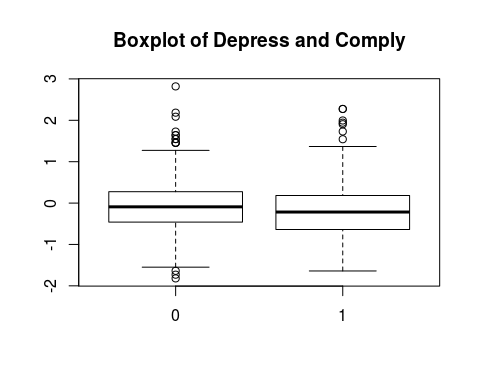
##### B5. comply and work



## comply  
## work 0 1  
## 0 357 249  
## 1 170 123

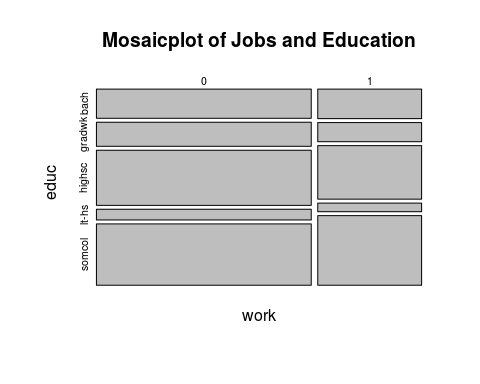
## x-squared DF p-value  
## 0.033065 1.000000 0.855700

##### B6 comply and depress



## t-statistics obs.diff in means p-value lower-bound of CI  
## 2.009100000 0.089312680 0.044870000 0.002047599  
## highr-bound of CI  
## 0.176577755

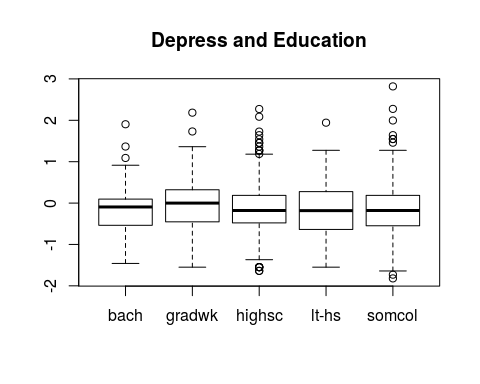
##### B7 education and work



## educ  
## work bach gradwk highsc lt-hs somcol  
## 0 98 81 185 36 206  
## 1 48 31 87 14 113

## x-squared DF p-value  
## 2.9255 4.0000 0.5704

##### B8 education and depress



## F-statistics p-value  
## 0.900 0.463

##### B9 age and depress



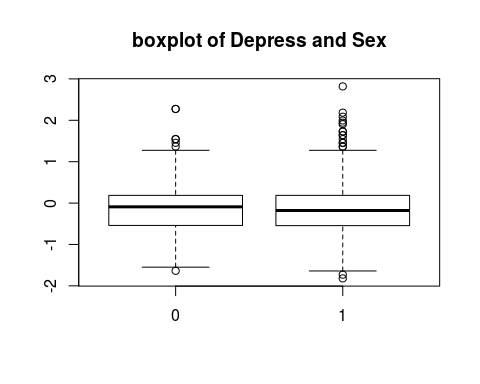
## estimates p-value  
## intercept -0.174786 0.031900  
## age 0.001229 0.556000

##### B10. age and work

## mean standard deviation sample size missing  
## un-reemployed 38.723 10.628 606.000 0.000  
## reemployed 35.169 9.663 293.000 0.000

## estimates p-value  
## intercept 5.38e-01 4.82e-02  
## age -3.42e-02 2.29e-06

##### B11. gender and depress



## mean standard deviation sample size missing  
## male -0.1269 0.6177 417.0000 0.0000  
## female -0.1301 0.6828 482.0000 0.0000

## t-statistics obs.diff in means p-value lower-bound of CI  
## 0.0722 0.0031 0.9424 -0.0820  
## highr-bound of CI  
## 0.0883

##### B12. gender and work

##   
## psyemp psyump  
## 0 0.1735261 0.2903226  
## 1 0.1523915 0.3837597

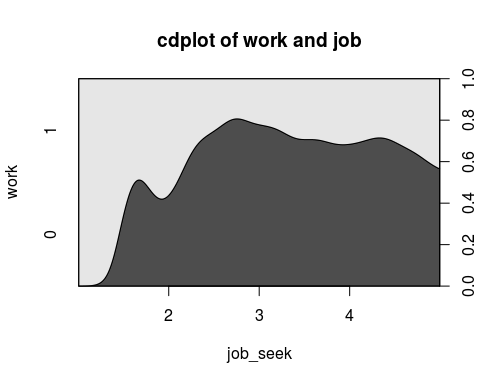
## chi-squared p-value  
## 7.815000 0.005182

##### B13. job\_seek and depress



## estimates p-value  
## intercept 0.2816 0.0214  
## job\_seek -0.1014 0.0007

##### B14. job\_seek and work



## mean standard deviation sample size missing  
## un-reemployed 4.0153 0.6834 606.0000 0.0000  
## reemployed 4.1010 0.8120 293.0000 0.0000

## estimates p-value  
## intercept -1.3983 0.0007  
## job\_seek 0.1654 0.0979

##### B15. chi-square test of work and comply

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table  
## X-squared = 0.033065, df = 1, p-value = 0.8557

##### B16. t-test of depress and comply

##   
## Welch Two Sample t-test  
##   
## data: depress by comply  
## t = 2.0091, df = 778.7, p-value = 0.04487  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.002047599 0.176577755  
## sample estimates:  
## mean in group 0 mean in group 1   
## -0.09167462 -0.18098730

##### B17. chi-square test of work and education

##   
## Pearson's Chi-squared test  
##   
## data: table3  
## X-squared = 2.9255, df = 4, p-value = 0.5704

##### B18. chi-square test of depress and education

## Warning in chisq.test(table4): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: table4  
## X-squared = 1763, df = 1732, p-value = 0.2963

##### B19. Anova for depress and occupation

## Df Sum Sq Mean Sq F value Pr(>F)  
## occp 6 4 0.6671 1.57 0.153  
## Residuals 892 379 0.4249

##### B20. Anova for depress and marital status

## Df Sum Sq Mean Sq F value Pr(>F)  
## marital 4 1.5 0.3804 0.891 0.468  
## Residuals 894 381.5 0.4267

##### B21. chi-square test for work and occupation

##   
## Pearson's Chi-squared test  
##   
## data: work.table  
## X-squared = 11.347, df = 6, p-value = 0.07822

##### B22. chi-square test for work and marital

##   
## Pearson's Chi-squared test  
##   
## data: mar.table  
## X-squared = 6.9897, df = 4, p-value = 0.1364

##### B23. linear regression for depress and age

##   
## Call:  
## lm(formula = depress ~ age, data = jobs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.71962 -0.40893 -0.00538 0.32635 2.94848   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.174786 0.081330 -2.149 0.0319 \*  
## age 0.001229 0.002086 0.589 0.5560   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6533 on 897 degrees of freedom  
## Multiple R-squared: 0.0003866, Adjusted R-squared: -0.0007278   
## F-statistic: 0.347 on 1 and 897 DF, p-value: 0.556

##### B24. logistic regression for work and age

##   
## Call:  
## glm(formula = work ~ age, family = binomial(link = "logit"),   
## data = jobs)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1515 -0.9233 -0.7842 1.3493 1.8992   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.538241 0.272499 1.975 0.0482 \*   
## age -0.034283 0.007254 -4.726 2.29e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1135.0 on 898 degrees of freedom  
## Residual deviance: 1111.4 on 897 degrees of freedom  
## AIC: 1115.4  
##   
## Number of Fisher Scoring iterations: 4

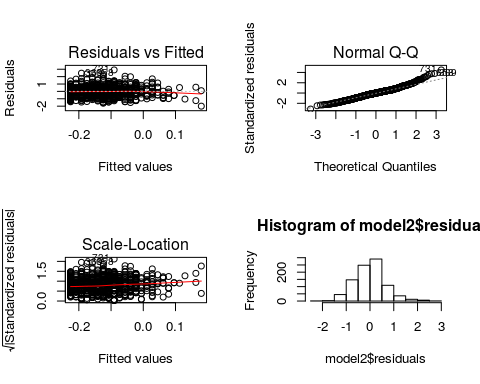
##### B25. t.test for depress and gender

##   
## Welch Two Sample t-test  
##   
## data: depress by sex  
## t = 0.072216, df = 895.2, p-value = 0.9424  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.08201479 0.08828099  
## sample estimates:  
## mean in group 0 mean in group 1   
## -0.1269518 -0.1300849

##### B26. chi-squared test for work and gender

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table1  
## X-squared = 7.815, df = 1, p-value = 0.005182

##### B27. Diagnostic plot for depress and job\_seek



##### B28. linear regression for depress and job\_seek

##   
## Call:  
## lm(formula = depress ~ job\_seek, data = jobs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.0001 -0.4112 0.0055 0.3447 2.9095   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.28157 0.12219 2.304 0.021432 \*   
## job\_seek -0.10145 0.02974 -3.411 0.000676 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6492 on 897 degrees of freedom  
## Multiple R-squared: 0.01281, Adjusted R-squared: 0.0117   
## F-statistic: 11.64 on 1 and 897 DF, p-value: 0.0006761

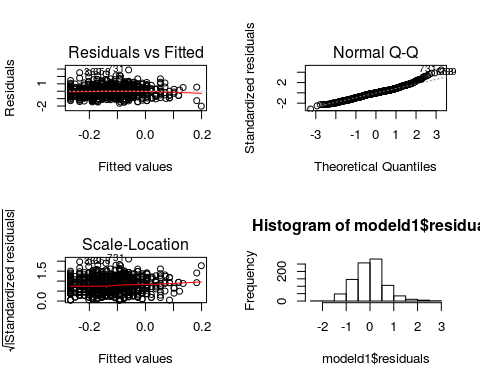
## 2.5 % 97.5 %  
## (Intercept) 0.04175434 0.5213768  
## job\_seek -0.15982144 -0.0430785

##### B29. logistic regression for work and job\_seek

##   
## Call:  
## glm(formula = work ~ job\_seek, family = "binomial", data = jobs)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9464 -0.9049 -0.8645 1.4399 1.7255   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.39827 0.41351 -3.382 0.000721 \*\*\*  
## job\_seek 0.16545 0.09996 1.655 0.097880 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1135.0 on 898 degrees of freedom  
## Residual deviance: 1132.2 on 897 degrees of freedom  
## AIC: 1136.2  
##   
## Number of Fisher Scoring iterations: 4

#### Appendix C: Multiple linear regression:

##### C1. Diagnostic plot for depress=comply+job\_seek



##### C2. depress=comply + job\_seek

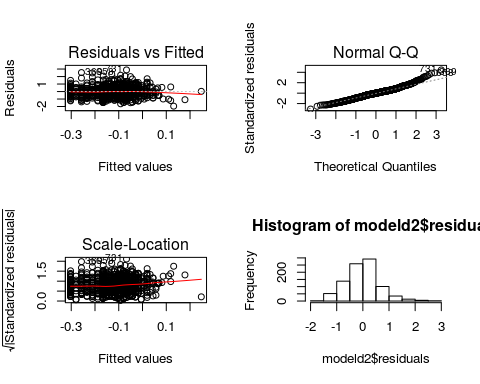
##   
## Call:  
## lm(formula = depress ~ comply + job\_seek, data = jobs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.0191 -0.4198 -0.0031 0.3534 2.8791   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.29620 0.12233 2.421 0.01566 \*   
## comply -0.07750 0.04406 -1.759 0.07892 .   
## job\_seek -0.09714 0.02981 -3.259 0.00116 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6485 on 896 degrees of freedom  
## Multiple R-squared: 0.0162, Adjusted R-squared: 0.01401   
## F-statistic: 7.378 on 2 and 896 DF, p-value: 0.0006634

## 2.5 % 97.5 %  
## (Intercept) 0.05611361 0.536290058  
## comply -0.16398081 0.008972318  
## job\_seek -0.15563978 -0.038636372

##### C3. Nested F test for variable comply

## Analysis of Variance Table  
##   
## Model 1: depress ~ job\_seek  
## Model 2: depress ~ comply + job\_seek  
## Res.Df RSS Df Sum of Sq Pr(>Chi)   
## 1 897 378.08   
## 2 896 376.77 1 1.3011 0.07858 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##### C4. Diagnostic plot for depress=comply + job\_seek + comply\*job\_seek



##### C5. depress=comply + job\_seek + comply \* job\_seek

##   
## Call:  
## lm(formula = depress ~ comply + job\_seek + comply \* job\_seek,   
## data = jobs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.95112 -0.41591 -0.00326 0.34778 2.88647   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.20556 0.15140 1.358 0.175   
## comply 0.17941 0.25665 0.699 0.485   
## job\_seek -0.07444 0.03725 -1.998 0.046 \*  
## comply:job\_seek -0.06311 0.06211 -1.016 0.310   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6485 on 895 degrees of freedom  
## Multiple R-squared: 0.01734, Adjusted R-squared: 0.01404   
## F-statistic: 5.263 on 3 and 895 DF, p-value: 0.001331

## 2.5 % 97.5 %  
## (Intercept) -0.09159024 0.502700306  
## comply -0.32430134 0.683121707  
## job\_seek -0.14754604 -0.001327741  
## comply:job\_seek -0.18500861 0.058787979

##### C6. Backward selection procedure

## Start: AIC=-774.84  
## depress ~ comply + job\_seek + comply \* job\_seek  
##   
## Df Sum of Sq RSS AIC  
## - comply:job\_seek 1 0.43414 376.77 -775.80  
## <none> 376.34 -774.84  
##   
## Step: AIC=-775.8  
## depress ~ comply + job\_seek  
##   
## Df Sum of Sq RSS AIC  
## <none> 376.77 -775.80  
## - comply 1 1.3011 378.08 -774.70  
## - job\_seek 1 4.4657 381.24 -767.21

##   
## Call:  
## lm(formula = depress ~ comply + job\_seek, data = jobs)  
##   
## Coefficients:  
## (Intercept) comply job\_seek   
## 0.29620 -0.07750 -0.09714

#### Appendix D: Multiple logistic regression:

##### D1. work=age + sex

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 1.2039388 3.6590621  
## age 0.9526203 0.9802398  
## sex 0.5039444 0.8892654

##   
## Call:  
## glm(formula = work ~ age + sex, family = "binomial", data = jobs)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1905 -0.9191 -0.7584 1.3314 1.9603   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.737832 0.283338 2.604 0.00921 \*\*   
## age -0.034066 0.007284 -4.676 2.92e-06 \*\*\*  
## sex -0.400745 0.144799 -2.768 0.00565 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1135.0 on 898 degrees of freedom  
## Residual deviance: 1103.7 on 896 degrees of freedom  
## AIC: 1109.7  
##   
## Number of Fisher Scoring iterations: 4

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 1.006599 2.9319057  
## age 0.952470 0.9799669

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.4892718 0.7277466  
## sex 0.5016542 0.8788744

##   
## Call:  
## glm(formula = work ~ sex, family = "binomial", data = jobs)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9681 -0.9681 -0.8178 1.4023 1.5862   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.5147 0.1012 -5.086 3.67e-07 \*\*\*  
## sex -0.4089 0.1430 -2.860 0.00424 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1135.0 on 898 degrees of freedom  
## Residual deviance: 1126.8 on 897 degrees of freedom  
## AIC: 1130.8  
##   
## Number of Fisher Scoring iterations: 4

##### D2. work=age + sex + age \* sex

##   
## Call:  
## glm(formula = work ~ age + sex + age \* sex, family = "binomial",   
## data = jobs)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1685 -0.9262 -0.7591 1.3322 1.9947   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.617264 0.392958 1.571 0.11623   
## age -0.030745 0.010435 -2.946 0.00321 \*\*  
## sex -0.168156 0.547138 -0.307 0.75859   
## age:sex -0.006422 0.014572 -0.441 0.65943   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1135.0 on 898 degrees of freedom  
## Residual deviance: 1103.5 on 895 degrees of freedom  
## AIC: 1111.5  
##   
## Number of Fisher Scoring iterations: 4

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.8631742 4.0387790  
## age 0.9497036 0.9894364  
## sex 0.2884685 2.4683081  
## age:sex 0.9655999 1.0224362

##### D3. Backward selection procedure

## Start: AIC=1111.55  
## work ~ age + sex + age \* sex  
##   
## Df Deviance AIC  
## - age:sex 1 1103.7 1109.7  
## <none> 1103.5 1111.5  
##   
## Step: AIC=1109.74  
## work ~ age + sex  
##   
## Df Deviance AIC  
## <none> 1103.7 1109.7  
## - sex 1 1111.4 1115.4  
## - age 1 1126.8 1130.8

##   
## Call: glm(formula = work ~ age + sex, family = "binomial", data = jobs)  
##   
## Coefficients:  
## (Intercept) age sex   
## 0.73783 -0.03407 -0.40075   
##   
## Degrees of Freedom: 898 Total (i.e. Null); 896 Residual  
## Null Deviance: 1135   
## Residual Deviance: 1104 AIC: 1110

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: work.table  
## X-squared = 7.815, df = 1, p-value = 0.005182

##### D4. Conditional Density Plot for Age and Work

**cdplot**(work~age, main="cdplot of work and age", data=jobs)

