PSTAT 175 Final Project

David Li, Aryanka Thaker, Sarah Liang

2022-12-12

Introduction

Goal of this project

The goal of this project is to predict the employee turnover time using the survival analysis method. In this dataset, we have 16 columns, with **stag** as the months that a employee quit and **event** as the censoring status. Besides interest in the impact of independent variables like **age** and **industry** on our survival function, we are more interested in the peculiar variables like **extraversion**, **independ**, **selfcontrol**, **novator**, and anxiety which can be indicative towards the employee's personality traits. We are wondering: is there an efficient way to manipulate the personality variables such that we are able to determine if there is a particular combination of personality traits with significant effect on employee turnover? From our project, it would be interesting to find out the type of employee most likely to stay with employers.

Data Description

There are 1129 observations and 16 columns in this data set. We have 14 independent variables:

- gender - employee's gender (male or female)
- age - employee's age in years
- industry - the industry in which the employee works
- profession - occupation of the department under which the employee's occupation lies
- traffic - how the employee found the company
- coach - whether or not there was a training coach present during the probationary period
- head_gender - gender (male or female) of the employee's supervisor
- greywage - whether the salary was minimum wage (grey) or just above minimum (white)
- way - employee's mode of transportation
- anxiety - employee's anxiety score

Four personality Traits: – extraversion - employee's extroversion score (0-10) – independ - employee's independence score (0-10) – selfcontrol - employee's self-control score (0-10) – novator - employee's innovation score (0-10)

Reference

Data obtained from Kaggle: Employee Turnover, uploaded by DAVIN WIJAYA URL: https://www.kaggle.com/datasets/davinwijaya/employee-turnover

Step Function R document:

Hastie, T. J. and Pregibon, D. (1992) Generalized linear models. Chapter 6 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. New York: Springer (4th ed).

Survival Analysis Functions:

Therneau, Terry. A package for Survival analysis in R

Carter, Andrew. Lecture slides for Fall 2022 at UC Santa Barbara

Survival Analysis

Import data & Library

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.2
```

```
## Warning: package 'tibble' was built under R version 4.1.2
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
## Warning: package 'forcats' was built under R version 4.1.2
```

```
library(ggplot2)
library(survival)
library(survminer)
library(janitor)
library(formatR)
```

```
## Warning: package 'formatR' was built under R version 4.1.2
```

```
turnover = read.csv("Turnover.csv")
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.1.2
```

```
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

Cleaning

```
turnover = turnover %>%
    clean_names(parsing_option = 0)
turnover$profession[which(turnover$profession == "Finan\xf1e")] = "Finance"
```

Assumptions

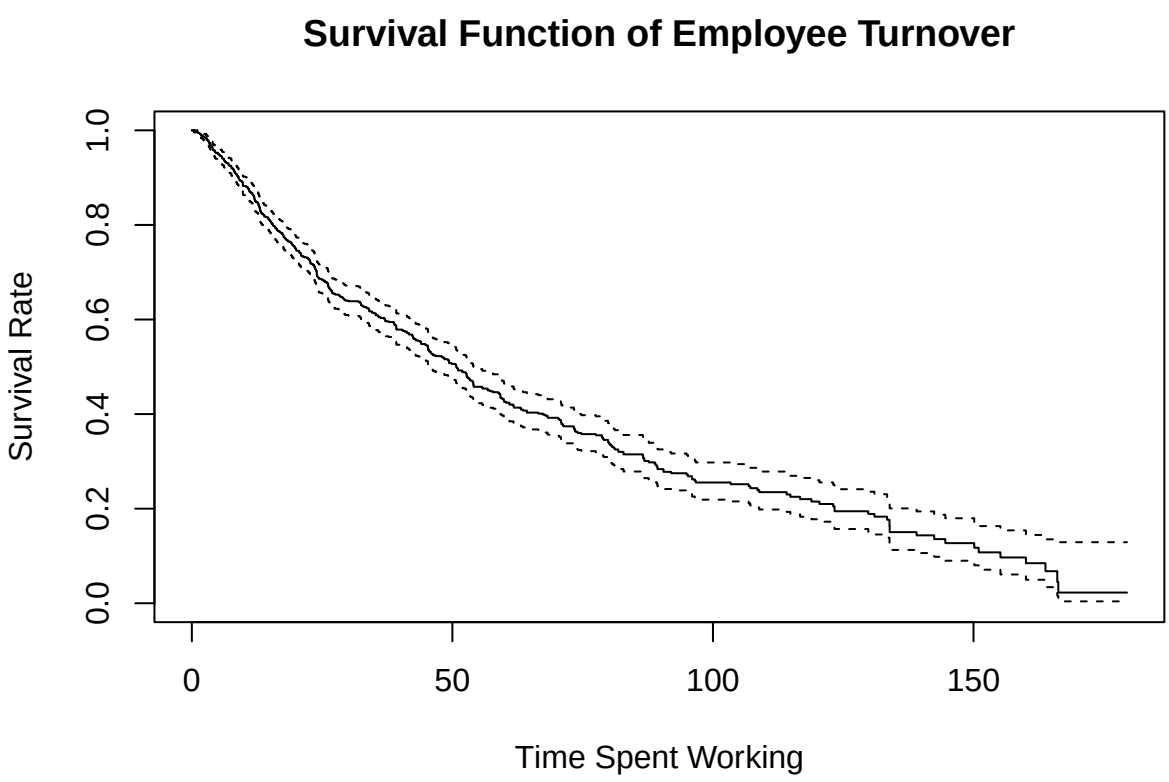
1. The optimal model is selected based on AIC criterion.
2. The model satisfies proportional hazard assumptions.
3. The variables and the model are significant at 95% confidence level.
4. The variables are not time dependent. Among these, age is a time dependent variable. To remove the time dependency, we create a new variable ageentry to indicate the age that the candidate entered the study.

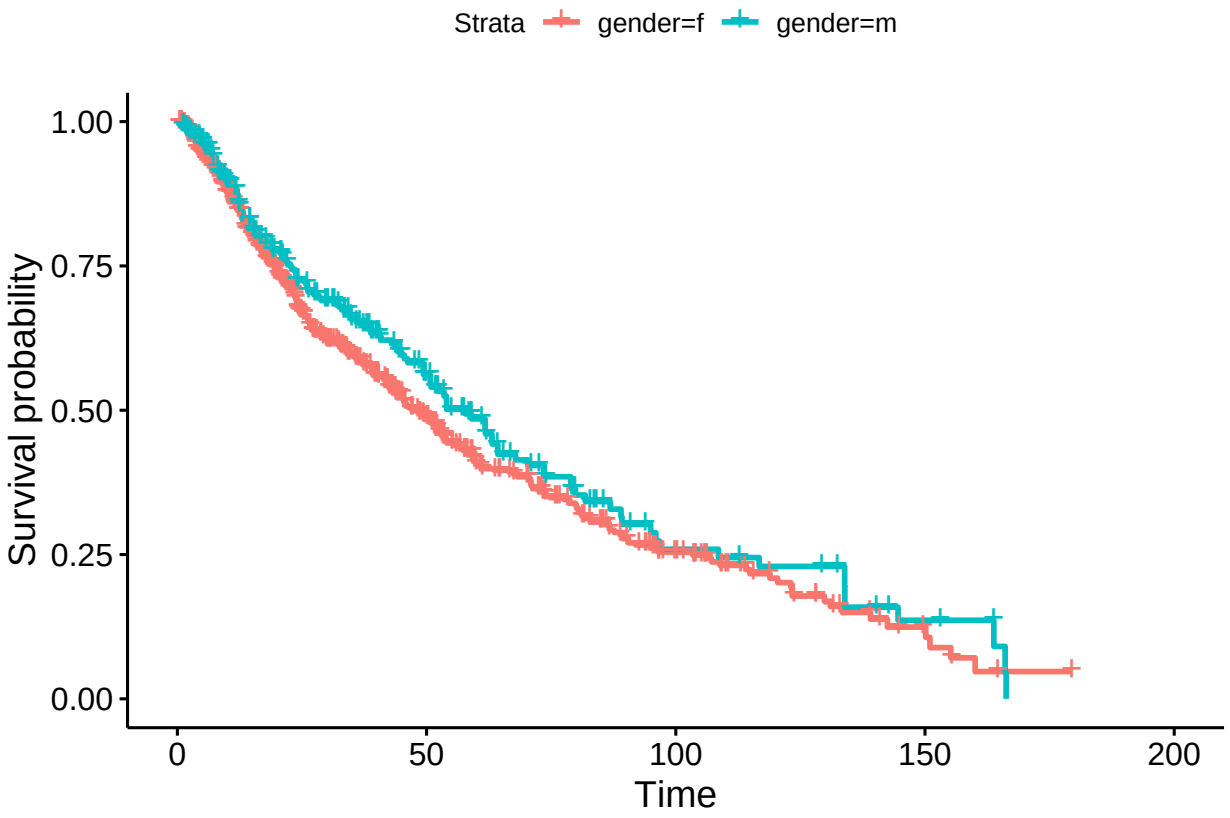
```
turnover = turnover %>%
    mutate(ageentry = age - stag/12) %>%
    select(-age)
```

Kaplan Meier Estimate

With a basic Kaplan-Meier Estimate, we can see the data follows a nice survival curve. We should be able to work with this data nicely

```
turnover.km <- survfit(Surv(stag, event) ~ 1, data = turnover)
plot(turnover.km, xlab = "Time Spent Working", ylab = "Survival Rate",
    main = "Survival Function of Employee Turnover")
ggsurvplot(survfit(Surv(stag, event) ~ gender, data = turnover))
```


Survival Function of Employee Turnover
Survival Rate
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
Time Spent Working




```
model <- coxph(Surv(stag, event) ~ gender, data = turnover)
```

We can begin to look at the specifics of the data: females and males seem to follow a similar pattern in turnover survival rate, but females tend to have a lower survival rate. This means females may quit faster than males, however the difference between the two genders is not significant, with a p-value of 0.1262732 > 0.05.

This method is straight forward, but we want to eliminate the effect of confounding variables and build one cox proportional model with all significant variables in it.

Cox Proportional Model

Here we are investigating the effect of several variables (both general statistics about the subjects as well as personality traits) from the employee turnover data upon the time the event took place. We are using **forward selection** that adds one variable to minimize AIC for each iteration, until adding more variables will not reduce AIC.

We use the *step* function to do the iteration automatically. The process is not printed.

```
turnover.null <- coxph(Surv(stag, event) ~ 1, data = turnover)
turnover.full <- coxph(Surv(stag, event) ~ ., data = turnover)
# iterate for optimal model that minimizes AIC
step(turnover.null, scope = list(lower = turnover.null, upper = turnover.full),
    direction = "forward")
```

Using the forward selection, we have the following variables that are significant with an AIC of 6737.634. Among the personality traits, only **extraversion** is significant.

```
model.1 <- coxph(formula = Surv(stag, event) ~ ageentry + industry +
    extraversion + greywage + traffic + coach + way + profession +
    anxiety, data = turnover)
anova(model.1)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                loglik    Chisq Df Pr(>|Chi|)
## NULL          -3470.3
## ageentry      -3399.9 140.7395  1  < 2.2e-16 ***
## industry      -3370.2  59.4176 15  3.175e-07 ***
## extraversion  -3364.3  11.7778  1  0.0005994 ***
## greywage      -3358.0  12.6815  1  0.0003693 ***
## traffic       -3347.0  21.9349  7  0.0026069 **
## coach         -3343.8   6.3899  2  0.0409694 *
## way           -3340.5   6.6254  2  0.0364170 *
## profession    -3326.0  29.0675 14  0.0102328 *
## anxiety       -3324.8   2.3337  1  0.1265972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(model.1)
```

```
## [1] 6737.634
```

Validating Assumptions & Stratified Model

We use the cox.zph function to test the proportional hazards assumption of our previous cox regression. From this test we can see that the variable profession **violates** the proportional hazard assumption due to its p-value of 0.00048 which is less than the 0.05 significance value.

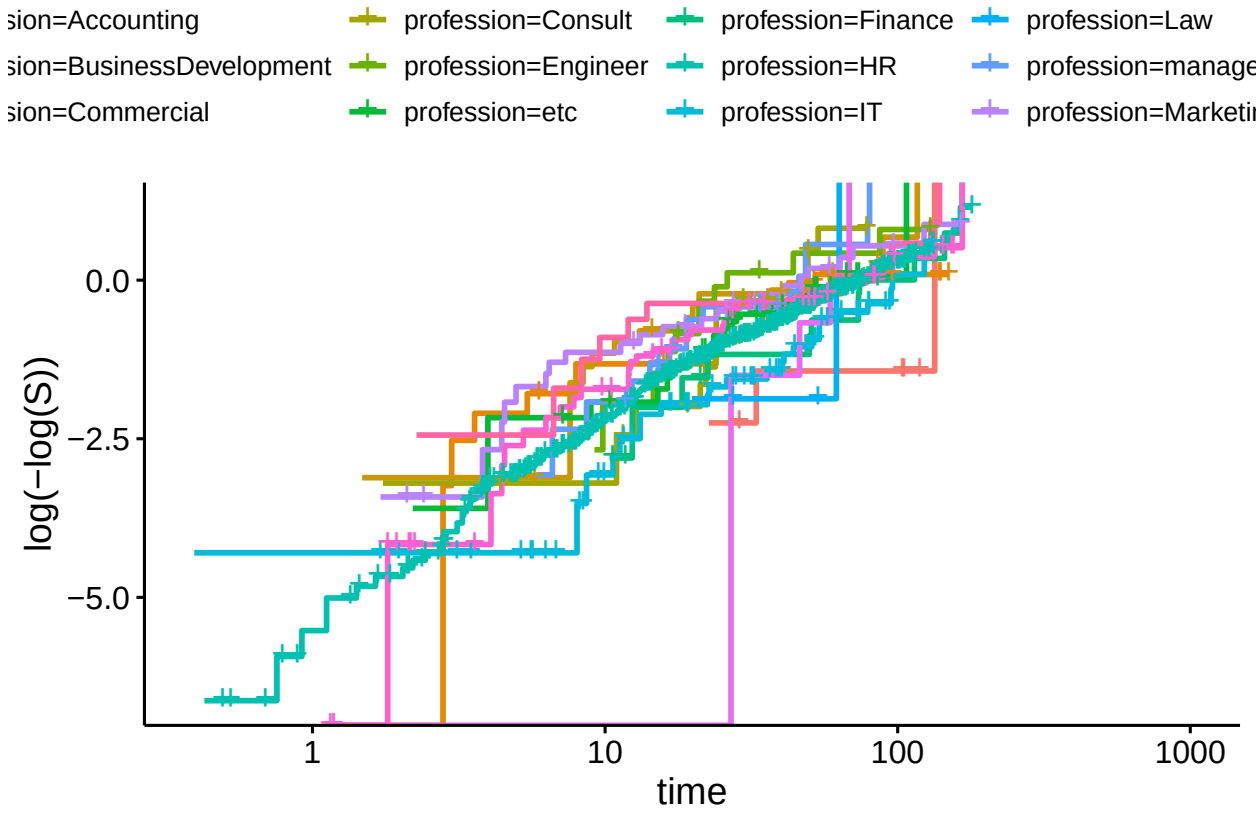
The log-log plot is not a good way to see, but we can see some lines crossing each other, which supports our argument.

```
cox.zph(model.1)
```

```
##                 chisq df       p
## ageentry     1.51e-04  1 0.99020
## industry     1.62e+01 15 0.36747
## extraversion 1.44e+00  1 0.22938
## greywage     3.64e-01  1 0.54605
## traffic      1.21e+01  7 0.09744
## coach        1.83e+00  2 0.40148
## way          1.58e+00  2 0.45401
## profession   4.14e+01 14 0.00015
## anxiety      1.85e+00  1 0.17417
## GLOBAL       8.09e+01 44 0.00059
```

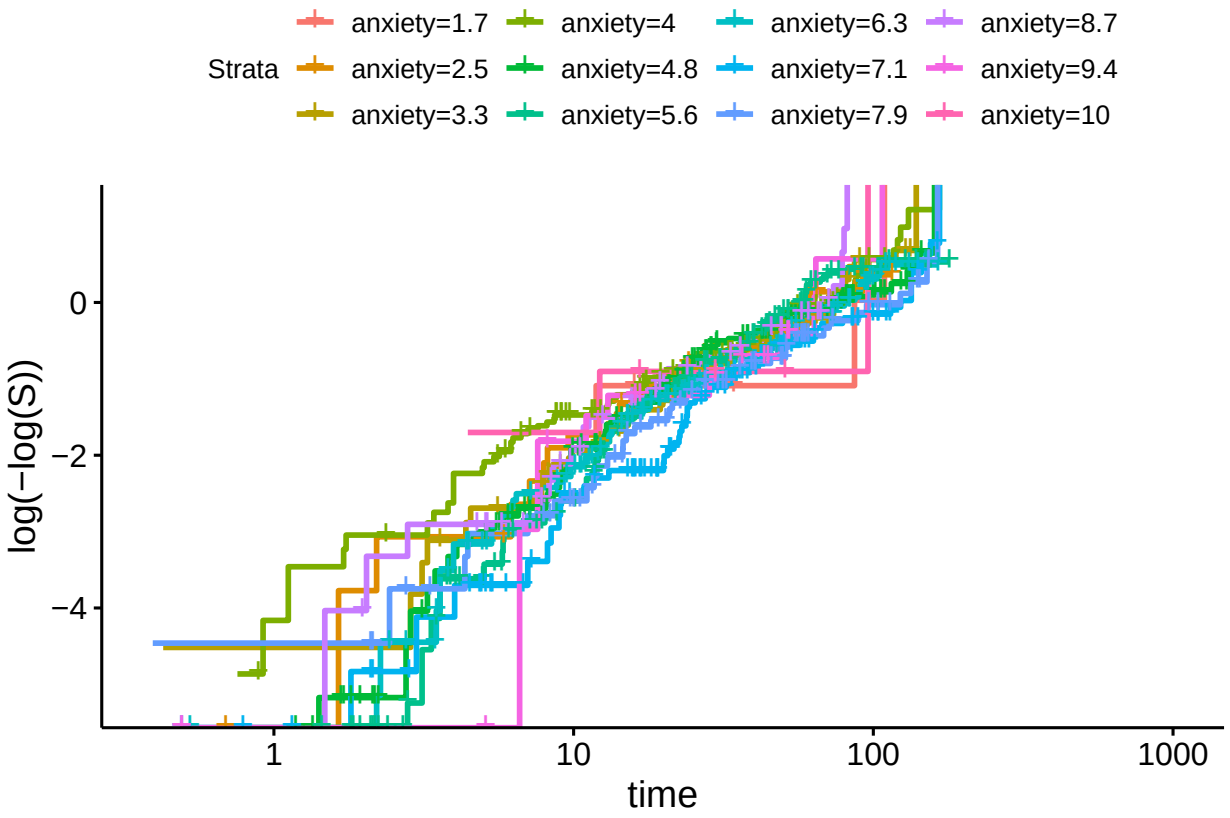
```
cloglog <- function(x) {
    log(-log(x))
}
ggsurvplot(survfit(Surv(stag, event) ~ profession, data = turnover),
    fun = "cloglog", xlab = "time", ylab = "log(-log(S))", main = "C-loglog Plot for Profession
Variable",
    pch = rep(19, 2))
```

```
## Warning: Ignoring unknown parameters: shape
```



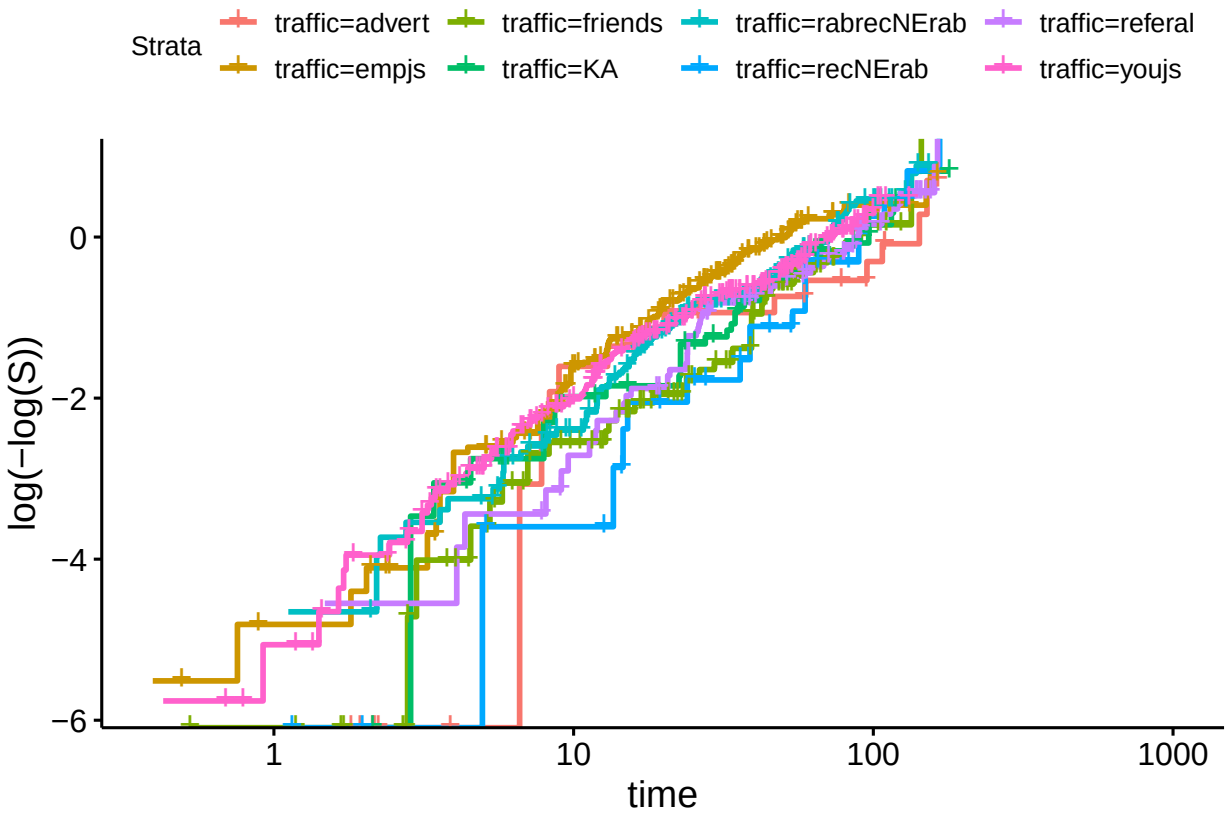

```
ggsurvplot(survfit(Surv(stag, event) ~ anxiety, data = turnover),
    fun = "cloglog", xlab = "time", ylab = "log(-log(S))", main = "C-loglog Plot for Anxiety
     Variable",
    pch = rep(19, 2))
```

```
## Warning: Ignoring unknown parameters: shape
```




```
ggsurvplot(survfit(Surv(stag, event) ~ traffic, data = turnover),
    fun = "cloglog", xlab = "time", ylab = "log(-log(S))", main = "C-loglog Plot for Traffic
     Variable",
    pch = rep(19, 2))
```

```
## Warning: Ignoring unknown parameters: shape
```



To solve this issue, we stratify the data by profession. This will divide the data proportional to the professions and fit a new model. It turns out that the anova table does not look very different than before, and we have a new AIC level of 4961.391.

```
model.2 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + anxiety, data = turnover)
anova(model.2)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                loglik    Chisq Df Pr(>|Chi|)
## NULL          -2571.6
## ageentry      -2506.0 131.2520  1  < 2.2e-16 ***
## industry      -2480.6  50.7195 15   9.18e-06 ***
## extraversion -2475.7   9.8195  1  0.0017267 **
## greywage      -2471.1   9.1890  1  0.0024348 **
## traffic       -2458.3  25.5721  7  0.0006004 ***
## coach         -2454.5   7.7050  2  0.0212267 *
## way           -2451.8   5.4444  2  0.0657291 .
## anxiety       -2450.7   2.1365  1  0.1438308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(model.2)  #4961.391
```

```
## [1] 4961.391
```

Conclusion on Hazard Rates

While we have found plenty variables that affect the turnover time significantly, we are more interested in analyzing the effect of **personality traits** and **anxiety**.

From reading the hazard ratio (exp(coef)) and it's confidence interval, we discover that a higher extraversion increases the turnover rate over time (hazard ratio > 1), while a higher anxiety decreases the turnover rate over time (hazard ratio < 1).

The result is counter intuitive for anxiety since people would be more likely to change their job if they are more anxious. However, it's important to note that the confidence interval for anxiety is not strictly below 1. In this case, it's likely that the story is other way around. The confidence interval for extraversion is strictly above 1.

```
summary(model.2)$conf.int
```

```
##                           exp(coef) exp(-coef) lower .95 upper .95
## ageentry                  1.0734296  0.9315934 1.0607395 1.0862716
## industryAgriculture       2.3443958  0.4265491 0.8074620 6.8067495
## industryBanks             1.7357429  0.5761222 0.7363786 4.0913785
## industryBuilding          1.6385172  0.6103079 0.6663790 4.0288461
## industryConsult           1.4072557  0.7106029 0.5815800 3.4051525
## industryetc               1.1244178  0.8893492 0.4701997 2.6888904
## industryIT                0.6387662  1.5655181 0.2614796 1.5604359
## industrymanufacture       0.7779509  1.2854281 0.3293253 1.8377202
## industryMining            0.9753845  1.0252367 0.3419908 2.7818728
## industryPharma            1.0513514  0.9511568 0.3734055 2.9601594
## industryPowerGeneration 0.8830873  1.1323909 0.3336197 2.3375215
## industryRealEstate        0.5659684  1.7668831 0.1663709 1.9253380
## industryRetail            0.8953074  1.1169348 0.3865621 2.0736002
## industryState             1.0200862  0.9803093 0.3971054 2.6204025
## industryTelecom           0.7904360  1.2651246 0.2975172 2.1000097
## industrytransport         0.8902791  1.1232432 0.3411615 2.3232312
## extraversion              1.0733283  0.9316814 1.0215873 1.1276899
## greywagewhite             0.6733821  1.4850410 0.5160463 0.8786875
## trafficempjs              1.6750342  0.5970027 0.8937022 3.1394571
## trafficfriends            0.8371332  1.1945531 0.4236677 1.6541073
## trafficKA                 0.9678705  1.0331961 0.4779328 1.9600524
## trafficrabrecNErab        1.4712820  0.6796794 0.7879798 2.7471146
## trafficrecNErab           0.9450024  1.0581984 0.4407172 2.0263099
## trafficreferal            1.3288242  0.7525450 0.6893823 2.5613855
## trafficyoujs              1.4698461  0.6803433 0.7871281 2.7447216
## coachno                   0.8886921  1.1252491 0.7143299 1.1056147
## coachyes                  1.2941508  0.7727074 0.9606352 1.7434573
## waycar                    0.8723584  1.1463178 0.7136621 1.0663439
## wayfoot                   0.7063843  1.4156600 0.4964392 1.0051156
## anxiety                   0.9615036  1.0400378 0.9121157 1.0135656
```

Advanced modeling

Methods

This section explores the non-linear relationship between the turnover time and the variables.

From the previous section, we found out some variables that are significant. Among the personality variables, only **extraversion** is significant. However, it's possible that other personality variables are correlated with the time in a non-linear relationship.

Indicator Transformation

To see the possible correlation, we create new variables to transform the personality scores into binary indicators. If the value is greater than 7, the indicator equals 1; If the value is less or equal to 7, the indicator equals 0. This transformation will magnify the difference between high value of personality traits and the lower ones.

Square Transformation

To see whether a personality traits has more effect when there are at extremes (0, 10), we takes the variable minus 5.5 and square it. If the variable is significant, it means the values at extremes has significant effect than the middle values on the time.

Coding

```
turnover_new = turnover %>%
    mutate(Great_personality = (extraversion + independ + selfcontrol +
        novator) > 25) %>%
    mutate(Great_personality = as.numeric(Great_personality)) %>%
    mutate(anxiety_square = (anxiety - 5)^2) %>%
    mutate(anxiety_indicator = as.numeric(anxiety > 7)) %>%
    mutate(selfcontrol_square = (selfcontrol - 5)^2) %>%
    mutate(selfcontrol_indicator = as.numeric(selfcontrol > 7)) %>%
    mutate(extraversion_indicator = as.numeric(extraversion >
        7)) %>%
    mutate(extraversion_square = (extraversion - 5)^2) %>%
    mutate(independ_indicator = as.numeric(independ > 7)) %>%
    mutate(independ_square = (independ - 5)^2) %>%
    mutate(novator_indicator = as.numeric(novator > 7)) %>%
    mutate(novator_square = (novator - 5)^2)
```

```
turnover.null <- coxph(Surv(stag, event) ~ 1, data = turnover_new)
turnover.full <- coxph(Surv(stag, event) ~ ., data = turnover_new)
step(turnover.null, scope = list(lower = turnover.null, upper = turnover.full),
    direction = "forward")
```

```
model.innovation = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + anxiety_indicator + independ_indicator + selfcontrol_square,
    data = turnover_new)
```

Analysis on Model Results

For the innovation model, in addition to the variables that are previously significant, we have discovered the significant effect of **independ_indicator** and **selfcontrol_square**. Also, **anxiety** changes to **anxiety_indicator**.

Comparing to the model from the previous part, the new model lowers the AIC from 4961.391 to 4955.029, which indicates an improvement.

```
AIC(model.2)  #4961.391
```

```
## [1] 4961.391
```

```
AIC(model.innovation)  #4955.029
```

```
## [1] 4955.725
```

Checking the test statistics of the model variables, **anxiety_indicator** shows an improvement of significance comparing to anxiety. **Independ_indicator** and **selfcontrol_square** are significant at 0.05 confidence level.

```
innovation.summary = summary(model.innovation)
anova(model.innovation)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                      loglik    Chisq Df Pr(>|Chi|)
## NULL                -2571.6
## ageentry            -2506.0 131.2520  1  < 2.2e-16 ***
## industry            -2480.6  50.7195 15   9.18e-06 ***
## extraversion        -2475.7   9.8195  1  0.0017267 **
## greywage            -2471.1   9.1890  1  0.0024348 **
## traffic             -2458.3  25.5721  7  0.0006004 ***
## coach               -2454.5   7.7050  2  0.0212267 *
## way                 -2451.8   5.4444  2  0.0657291 .
## anxiety_indicator   -2450.1   3.2940  1  0.0695315 .
## independ_indicator  -2448.1   4.0933  1  0.0430532 *
## selfcontrol_square  -2445.9   4.4158  1  0.0356077 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We only need to look at the last three lines, which are the confidence intervals for three personality traits. The upper bond of the hazard ratio for three variables are strictly below 1. It means that the higher of the rate, the slower that the candidate quits their job.

```
innovation.summary$conf.int
```

```
##                         exp(coef) exp(-coef) lower .95 upper .95
## ageentry                1.0731137  0.9318677 1.0602392 1.0861445
```

```
## industryAgriculture     2.1127938  0.4733070 0.7267561 6.1422219
## industryBanks           1.6907551  0.5914517 0.7181897 3.9803586
## industryBuilding        1.4886513  0.6717490 0.6050609 3.6625776
## industryConsult         1.3248068  0.7548270 0.5487872 3.1981669
## industryetc             1.0413228  0.9603170 0.4365248 2.4840582
## industryIT              0.5961953  1.6773026 0.2446044 1.4531584
## industrymanufacture     0.7441808  1.3437594 0.3147959 1.7592513
## industryMining          0.8929218  1.1199189 0.3135624 2.5427452
## industryPharma          0.9894800  1.0106318 0.3493593 2.8024754
## industryPowerGeneration 0.7723212  1.2947981 0.2913887 2.0470254
## industryRealEstate      0.5462239  1.8307510 0.1602230 1.8621587
## industryRetail          0.8430112  1.1862238 0.3636619 1.9541999
## industryState           0.8893435  1.1244249 0.3456453 2.2882758
## industryTelecom         0.7431942  1.3455433 0.2800804 1.9720678
## industrytransport       0.7910219  1.2641876 0.3025137 2.0683874
## extraversion            1.0481554  0.9540570 0.9946927 1.1044917
## greywagewhite           0.6849766  1.4599039 0.5247243 0.8941703
## trafficempjs            1.8756818  0.5331395 0.9934895 3.5412372
## trafficfriends          0.9257603  1.0801933 0.4647871 1.8439239
## trafficKA               1.0613247  0.9422187 0.5226421 2.1552230
## trafficrabrecNErab      1.6015984  0.6243762 0.8533972 3.0057719
## trafficrecNErab         1.0065000  0.9935420 0.4678113 2.1654931
## trafficreferal          1.4729353  0.6789165 0.7617665 2.8480357
## trafficyoujs            1.5997998  0.6250782 0.8525645 3.0019539
## coachno                 0.8823648  1.1333181 0.7094822 1.0973745
## coachyes                1.2700691  0.7873588 0.9433636 1.7099190
## waycar                  0.8670317  1.1533604 0.7094915 1.0595532
## wayfoot                 0.7111938  1.4060865 0.4998599 1.0118767
## anxiety_indicator       0.7838646  1.2757305 0.6312585 0.9733631
## independ_indicator      0.7354290  1.3597505 0.5657233 0.9560431
## selfcontrol_square      0.9812447  1.0191138 0.9636974 0.9991115
```

Extroversion

The conclusion on the extroversion score is the same as before, but the confidence interval for **extraversion** under the new model is not strictly above 1.

Anxiety

For the **anxiety_indicator**, it shows that there is a significant difference for an anxiety level above and below 7. The higher the anxiety, the slower one would quit. Comparing to the previous result, the confidence interval is strictly below 1 under the new model.

```
model.test1 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + independ_indicator + selfcontrol_square + anxiety,
    data = turnover_new)
anova(model.test1)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
```

```
##
##                       loglik    Chisq Df Pr(>|Chi|)
## NULL                 -2571.6
## ageentry             -2506.0 131.2520  1  < 2.2e-16 ***
## industry             -2480.6  50.7195 15   9.18e-06 ***
## extraversion         -2475.7   9.8195  1  0.0017267 **
## greywage             -2471.1   9.1890  1  0.0024348 **
## traffic              -2458.3  25.5721  7  0.0006004 ***
## coach                -2454.5   7.7050  2  0.0212267 *
## way                  -2451.8   5.4444  2  0.0657291 .
## independ_indicator   -2450.4   2.7784  1  0.0955440 .
## selfcontrol_square   -2448.4   4.0109  1  0.0452061 *
## anxiety              -2446.0   4.8279  1  0.0280021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.innovation1 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + independ_indicator + selfcontrol_square + anxiety_indicator,
    data = turnover_new)
anova(model.innovation1)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                       loglik    Chisq Df Pr(>|Chi|)
## NULL                 -2571.6
## ageentry             -2506.0 131.2520  1  < 2.2e-16 ***
## industry             -2480.6  50.7195 15   9.18e-06 ***
## extraversion         -2475.7   9.8195  1  0.0017267 **
## greywage             -2471.1   9.1890  1  0.0024348 **
## traffic              -2458.3  25.5721  7  0.0006004 ***
## coach                -2454.5   7.7050  2  0.0212267 *
## way                  -2451.8   5.4444  2  0.0657291 .
## independ_indicator   -2450.4   2.7784  1  0.0955440 .
## selfcontrol_square   -2448.4   4.0109  1  0.0452061 *
## anxiety_indicator    -2445.9   5.0138  1  0.0251457 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Independence

While fitting in **independ** will not be significant, it's surprising to see that **independ_indicator** works very well. It shows that this variable does not have a linear relationship, which means a value of 1 has the same effect as 6 or 7 on time. It is also interesting to see that more independent person tends to quit more slowly.

```
model.test2 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + selfcontrol_square + anxiety_indicator + independ,
    data = turnover_new)
anova(model.test2)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                      loglik    Chisq Df Pr(>|Chi|)
## NULL                -2571.6
## ageentry            -2506.0 131.2520  1  < 2.2e-16 ***
## industry            -2480.6  50.7195 15   9.18e-06 ***
## extraversion        -2475.7   9.8195  1  0.0017267 **
## greywage            -2471.1   9.1890  1  0.0024348 **
## traffic             -2458.3  25.5721  7  0.0006004 ***
## coach               -2454.5   7.7050  2  0.0212267 *
## way                 -2451.8   5.4444  2  0.0657291 .
## selfcontrol_square -2450.3   2.8990  1  0.0886345 .
## anxiety_indicator  -2448.6   3.3803  1  0.0659787 .
## independ            -2448.5   0.3237  1  0.5694039
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.innovation2 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + selfcontrol_square + anxiety_indicator + independ_indicator,
    data = turnover_new)
anova(model.innovation2)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                      loglik    Chisq Df Pr(>|Chi|)
## NULL                -2571.6
## ageentry            -2506.0 131.2520  1  < 2.2e-16 ***
## industry            -2480.6  50.7195 15   9.18e-06 ***
## extraversion        -2475.7   9.8195  1  0.0017267 **
## greywage            -2471.1   9.1890  1  0.0024348 **
## traffic             -2458.3  25.5721  7  0.0006004 ***
## coach               -2454.5   7.7050  2  0.0212267 *
## way                 -2451.8   5.4444  2  0.0657291 .
## selfcontrol_square -2450.3   2.8990  1  0.0886345 .
## anxiety_indicator  -2448.6   3.3803  1  0.0659787 .
## independ_indicator -2445.9   5.5238  1  0.0187593 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Self-Control

From the confidence intervals table, we know that the hazard ratio for **selfcontrol_square** is less than 1. It's important to note that after this transformation, 0 & 10 will both transform to 25 and becomes the maximum. In this case, the statistic shows that candidates with more extreme self-control scores tend to leave slower.

From the anova tables below, the **selfcontrol** variable is close to be significant. It might imply that some observations with low self-control scores are not consistent with the linear relationship. In this case, **self-control_square** is a better fit.

```
model.test3 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + anxiety_indicator + independ_indicator + selfcontrol,
    data = turnover_new)
anova(model.test3)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                      loglik    Chisq Df Pr(>|Chi|)
## NULL                -2571.6
## ageentry            -2506.0 131.2520  1  < 2.2e-16 ***
## industry            -2480.6  50.7195 15   9.18e-06 ***
## extraversion        -2475.7   9.8195  1  0.0017267 **
## greywage            -2471.1   9.1890  1  0.0024348 **
## traffic             -2458.3  25.5721  7  0.0006004 ***
## coach               -2454.5   7.7050  2  0.0212267 *
## way                 -2451.8   5.4444  2  0.0657291 .
## anxiety_indicator   -2450.1   3.2940  1  0.0695315 .
## independ_indicator  -2448.1   4.0933  1  0.0430532 *
## selfcontrol         -2446.6   3.0104  1  0.0827317 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.innovation3 = coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + extraversion + greywage + traffic +
    coach + way + anxiety_indicator + independ_indicator + selfcontrol_square,
    data = turnover_new)
anova(model.innovation3)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                      loglik    Chisq Df Pr(>|Chi|)
## NULL                -2571.6
## ageentry            -2506.0 131.2520  1  < 2.2e-16 ***
## industry            -2480.6  50.7195 15   9.18e-06 ***
## extraversion        -2475.7   9.8195  1  0.0017267 **
## greywage            -2471.1   9.1890  1  0.0024348 **
## traffic             -2458.3  25.5721  7  0.0006004 ***
## coach               -2454.5   7.7050  2  0.0212267 *
## way                 -2451.8   5.4444  2  0.0657291 .
## anxiety_indicator   -2450.1   3.2940  1  0.0695315 .
## independ_indicator  -2448.1   4.0933  1  0.0430532 *
## selfcontrol_square  -2445.9   4.4158  1  0.0356077 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conlusion on Advanced Analysis

In the advanced modeling section, we identify two more personality traits that have a significant effect on turnover time after transformations, which are independence and self-control. Also, we improve the model result by transforming anxiety into an indicator variable. Among four personality traits, independence, self-control, extroversion, and innovation, only innovation exhibits a non-significant correlation with turnover time. Independence and self-control become significant after indicator transformation and square transformation respectively.

From analyzing the coefficients, we conclude that people quit jobs slower with a higher anxiety score, higher independent score, and more extreme (close to 0 or 10) self-control score. Also, people stay in a job shorter if they have a higher extroversion score.

Interaction Terms - among personality traits

```
model.int1 <- coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + greywage + traffic + coach + way +
    extraversion * anxiety_indicator * independ_indicator * selfcontrol_square,
    data = turnover_new)
anova(model.int1)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                                                                          loglik
## NULL                                                                    -2571.6
## ageentry                                                                -2506.0
## industry                                                                -2480.6
## greywage                                                                -2476.6
## traffic                                                                 -2463.8
## coach                                                                   -2459.2
## way                                                                     -2456.4
## extraversion                                                            -2451.8
## anxiety_indicator                                                       -2450.1
## independ_indicator                                                      -2448.1
## selfcontrol_square                                                      -2445.9
## extraversion:anxiety_indicator                                          -2445.8
## extraversion:independ_indicator                                         -2442.7
## anxiety_indicator:independ_indicator                                    -2442.7
## extraversion:selfcontrol_square                                         -2442.6
## anxiety_indicator:selfcontrol_square                                    -2442.6
## independ_indicator:selfcontrol_square                                   -2442.0
## extraversion:anxiety_indicator:independ_indicator                       -2436.7
## extraversion:anxiety_indicator:selfcontrol_square                       -2436.1
## extraversion:independ_indicator:selfcontrol_square                      -2436.1
## anxiety_indicator:independ_indicator:selfcontrol_square                 -2434.8
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square -2431.5
##                                                                            Chisq
## NULL
## ageentry                                                                131.2520
## industry                                                                 50.7195
```

```
## greywage                                                          8.1287
## traffic                                                          25.5253
## coach                                                             9.1225
## way                                                               5.6001
## extraversion                                                      9.3533
## anxiety_indicator                                                 3.2940
## independ_indicator                                                4.0933
## selfcontrol_square                                                4.4158
## extraversion:anxiety_indicator                                    0.1705
## extraversion:independ_indicator                                   6.0663
## anxiety_indicator:independ_indicator                              0.0156
## extraversion:selfcontrol_square                                   0.2088
## anxiety_indicator:selfcontrol_square                              0.1290
## independ_indicator:selfcontrol_square                             1.0661
## extraversion:anxiety_indicator:independ_indicator                10.7606
## extraversion:anxiety_indicator:selfcontrol_square                 1.0467
## extraversion:independ_indicator:selfcontrol_square                0.0781
## anxiety_indicator:independ_indicator:selfcontrol_square           2.5034
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  6.7048
##                                                                      Df
## NULL
## ageentry                                                              1
## industry                                                             15
## greywage                                                              1
## traffic                                                               7
## coach                                                                 2
## way                                                                   2
## extraversion                                                          1
## anxiety_indicator                                                     1
## independ_indicator                                                    1
## selfcontrol_square                                                    1
## extraversion:anxiety_indicator                                        1
## extraversion:independ_indicator                                       1
## anxiety_indicator:independ_indicator                                  1
## extraversion:selfcontrol_square                                       1
## anxiety_indicator:selfcontrol_square                                  1
## independ_indicator:selfcontrol_square                                 1
## extraversion:anxiety_indicator:independ_indicator                     1
## extraversion:anxiety_indicator:selfcontrol_square                     1
## extraversion:independ_indicator:selfcontrol_square                    1
## anxiety_indicator:independ_indicator:selfcontrol_square               1
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  1
##                                                                  Pr(>|Chi|)
## NULL
## ageentry                                                        < 2.2e-16
## industry                                                         9.18e-06
## greywage                                                        0.0043571
## traffic                                                         0.0006121
## coach                                                           0.0104489
## way                                                             0.0608062
## extraversion                                                    0.0022258
## anxiety_indicator                                               0.0695315
## independ_indicator                                              0.0430532
## selfcontrol_square                                              0.0356077
```

```
## extraversion:anxiety_indicator                                          0.6796636
## extraversion:independ_indicator                                         0.0137789
## anxiety_indicator:independ_indicator                                    0.9004600
## extraversion:selfcontrol_square                                         0.6477176
## anxiety_indicator:selfcontrol_square                                    0.7195205
## independ_indicator:selfcontrol_square                                   0.3018220
## extraversion:anxiety_indicator:independ_indicator                       0.0010369
## extraversion:anxiety_indicator:selfcontrol_square                       0.3062759
## extraversion:independ_indicator:selfcontrol_square                      0.7799006
## anxiety_indicator:independ_indicator:selfcontrol_square                 0.1136037
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  0.0096152
##
## NULL
## ageentry                                                               ***
## industry                                                               ***
## greywage                                                               **
## traffic                                                                ***
## coach                                                                  *
## way                                                                    .
## extraversion                                                           **
## anxiety_indicator                                                      .
## independ_indicator                                                     *
## selfcontrol_square                                                     *
## extraversion:anxiety_indicator
## extraversion:independ_indicator                                        *
## anxiety_indicator:independ_indicator
## extraversion:selfcontrol_square
## anxiety_indicator:selfcontrol_square
## independ_indicator:selfcontrol_square
## extraversion:anxiety_indicator:independ_indicator                      **
## extraversion:anxiety_indicator:selfcontrol_square
## extraversion:independ_indicator:selfcontrol_square
## anxiety_indicator:independ_indicator:selfcontrol_square
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can explore some interaction terms between the four personality traits that resulted from our previous model that managed to minimize AIC the most: `extraversion`, `anxiety_indicator`, `independ_indicator`, and `selfcontrol_square`. By doing so, we discover that the interaction terms `anxiety_indicator:independ_indicator`, `extraversion:anxiety_indicator:independ_indicator`, `anxiety_indicator:independ_indicator:selfcontrol_square`, and `extraversion:anxiety_indicator:independ_indicator:selfcontrol_square` all have an associated p-value less than 0.05, thus they are all significant.

```
model.int2 <- coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + greywage + traffic + coach + way +
    extraversion + anxiety_indicator + independ_indicator + selfcontrol_square +
    anxiety_indicator:independ_indicator + extraversion:anxiety_indicator:independ_indicator +
    anxiety_indicator:independ_indicator:selfcontrol_square +
    extraversion:anxiety_indicator:independ_indicator:selfcontrol_square,
    data = turnover_new)

anova(model.int1, model.int2)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(stag, event)
##  Model 1: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversi
##  Model 2: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversi
##    loglik  Chisq Df P(>|Chi|)
## 1 -2431.5
## 2 -2434.3 5.6515  7     0.581
```

However, comparing the model with interaction terms and the mdoel without results in a p-value of 0.581 > 0.04. So the model is not significantly different or better than the model without the interaction terms.

```
anova(model.int2)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                                                                  loglik
## NULL                                                            -2571.6
## ageentry                                                        -2506.0
## industry                                                        -2480.6
## greywage                                                        -2476.6
## traffic                                                         -2463.8
## coach                                                           -2459.2
## way                                                             -2456.4
## extraversion                                                    -2451.8
## anxiety_indicator                                               -2450.1
## independ_indicator                                              -2448.1
## selfcontrol_square                                              -2445.9
## anxiety_indicator:independ_indicator                            -2445.8
## extraversion:anxiety_indicator:independ_indicator               -2439.1
## anxiety_indicator:independ_indicator:selfcontrol_square         -2437.7
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square -2434.3
##                                                                    Chisq
## NULL
## ageentry                                                        131.2520
## industry                                                         50.7195
## greywage                                                          8.1287
## traffic                                                          25.5253
## coach                                                             9.1225
## way                                                               5.6001
## extraversion                                                      9.3533
## anxiety_indicator                                                 3.2940
## independ_indicator                                                4.0933
## selfcontrol_square                                                4.4158
## anxiety_indicator:independ_indicator                              0.0938
## extraversion:anxiety_indicator:independ_indicator                13.5159
## anxiety_indicator:independ_indicator:selfcontrol_square           2.6726
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  6.8160
##                                                                  Df
## NULL
## ageentry                                                          1
## industry                                                         15
```

```
## greywage                                                                1
## traffic                                                                 7
## coach                                                                   2
## way                                                                     2
## extraversion                                                            1
## anxiety_indicator                                                       1
## independ_indicator                                                      1
## selfcontrol_square                                                      1
## anxiety_indicator:independ_indicator                                    1
## extraversion:anxiety_indicator:independ_indicator                       1
## anxiety_indicator:independ_indicator:selfcontrol_square                 1
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  1
##                                                                      Pr(>|Chi|)
## NULL
## ageentry                                                              < 2.2e-16
## industry                                                               9.18e-06
## greywage                                                              0.0043571
## traffic                                                               0.0006121
## coach                                                                 0.0104489
## way                                                                   0.0608062
## extraversion                                                          0.0022258
## anxiety_indicator                                                     0.0695315
## independ_indicator                                                    0.0430532
## selfcontrol_square                                                    0.0356077
## anxiety_indicator:independ_indicator                                  0.7593692
## extraversion:anxiety_indicator:independ_indicator                     0.0002366
## anxiety_indicator:independ_indicator:selfcontrol_square               0.1020920
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  0.0090345
##
## NULL
## ageentry                                                             ***
## industry                                                             ***
## greywage                                                             **
## traffic                                                              ***
## coach                                                                *
## way                                                                  .
## extraversion                                                         **
## anxiety_indicator                                                    .
## independ_indicator                                                   *
## selfcontrol_square                                                   *
## anxiety_indicator:independ_indicator
## extraversion:anxiety_indicator:independ_indicator                    ***
## anxiety_indicator:independ_indicator:selfcontrol_square
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We check the model and find that the following terms are not significant: `anxiety_indicator:independ_indicator` and `anxiety_indicator:independ_indicator:selfcontrol_square` (with p-values 0.7593 and 0.1020 $>$ 0.05 respectively).

```
model.int3 <- coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + greywage + traffic + coach + way +
    extraversion + anxiety_indicator + independ_indicator + selfcontrol_square +
```

```
    extraversion:anxiety_indicator:independ_indicator + extraversion:anxiety_indicator:independ_indicat
    data = turnover_new)
anova(model.int3, model.int1)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(stag, event)
##  Model 1: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversi
##  Model 2: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversi
##    loglik  Chisq Df P(>|Chi|)
## 1 -2440.0
## 2 -2431.5 17.099  9   0.04718 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By removing non-significant terms, we can achieve a significantly different model (p-value $0.0472 < 0.05$).

```
anova(model.int3)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                                                                  loglik
## NULL                                                            -2571.6
## ageentry                                                        -2506.0
## industry                                                        -2480.6
## greywage                                                        -2476.6
## traffic                                                         -2463.8
## coach                                                           -2459.2
## way                                                             -2456.4
## extraversion                                                    -2451.8
## anxiety_indicator                                               -2450.1
## independ_indicator                                              -2448.1
## selfcontrol_square                                              -2445.9
## extraversion:anxiety_indicator:independ_indicator               -2445.6
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square -2440.0
##                                                                    Chisq
## NULL
## ageentry                                                        131.2520
## industry                                                         50.7195
## greywage                                                          8.1287
## traffic                                                          25.5253
## coach                                                             9.1225
## way                                                               5.6001
## extraversion                                                      9.3533
## anxiety_indicator                                                 3.2940
## independ_indicator                                                4.0933
## selfcontrol_square                                                4.4158
## extraversion:anxiety_indicator:independ_indicator                 0.4499
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square 11.2004
##                                                                  Df
## NULL
## ageentry                                                          1
```

```
## industry                                                                 15
## greywage                                                                  1
## traffic                                                                   7
## coach                                                                     2
## way                                                                       2
## extraversion                                                              1
## anxiety_indicator                                                         1
## independ_indicator                                                        1
## selfcontrol_square                                                        1
## extraversion:anxiety_indicator:independ_indicator                         1
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square      1
##                                                                  Pr(>|Chi|)
## NULL
## ageentry                                                          < 2.2e-16
## industry                                                           9.18e-06
## greywage                                                          0.0043571
## traffic                                                           0.0006121
## coach                                                             0.0104489
## way                                                               0.0608062
## extraversion                                                      0.0022258
## anxiety_indicator                                                 0.0695315
## independ_indicator                                                0.0430532
## selfcontrol_square                                                0.0356077
## extraversion:anxiety_indicator:independ_indicator                 0.5023686
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square 0.0008178
##
## NULL
## ageentry                                                                ***
## industry                                                                ***
## greywage                                                                **
## traffic                                                                 ***
## coach                                                                   *
## way                                                                     .
## extraversion                                                            **
## anxiety_indicator                                                       .
## independ_indicator                                                      *
## selfcontrol_square                                                      *
## extraversion:anxiety_indicator:independ_indicator
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.int4 <- coxph(formula = Surv(stag, event) ~ strata(profession) +
    ageentry + industry + greywage + traffic + coach + way +
    extraversion + anxiety_indicator + independ_indicator + selfcontrol_square +
    extraversion:anxiety_indicator:independ_indicator:selfcontrol_square,
    data = turnover_new)
anova(model.int4)  # all significant
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                                                                   loglik
```

```
## NULL                                                                    -2571.6
## ageentry                                                                -2506.0
## industry                                                                -2480.6
## greywage                                                                -2476.6
## traffic                                                                 -2463.8
## coach                                                                   -2459.2
## way                                                                     -2456.4
## extraversion                                                            -2451.8
## anxiety_indicator                                                       -2450.1
## independ_indicator                                                      -2448.1
## selfcontrol_square                                                      -2445.9
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square -2442.4
##                                                                            Chisq
## NULL
## ageentry                                                               131.2520
## industry                                                                50.7195
## greywage                                                                 8.1287
## traffic                                                                 25.5253
## coach                                                                    9.1225
## way                                                                      5.6001
## extraversion                                                             9.3533
## anxiety_indicator                                                        3.2940
## independ_indicator                                                       4.0933
## selfcontrol_square                                                       4.4158
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square   6.8589
##                                                                           Df
## NULL
## ageentry                                                                 1
## industry                                                                15
## greywage                                                                 1
## traffic                                                                  7
## coach                                                                    2
## way                                                                      2
## extraversion                                                             1
## anxiety_indicator                                                        1
## independ_indicator                                                       1
## selfcontrol_square                                                       1
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  1
##                                                                      Pr(>|Chi|)
## NULL
## ageentry                                                              < 2.2e-16
## industry                                                               9.18e-06
## greywage                                                              0.0043571
## traffic                                                               0.0006121
## coach                                                                 0.0104489
## way                                                                   0.0608062
## extraversion                                                          0.0022258
## anxiety_indicator                                                     0.0695315
## independ_indicator                                                    0.0430532
## selfcontrol_square                                                    0.0356077
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  0.0088203
##
## NULL
## ageentry                                                                 ***
```

```
## industry                                                                 ***
## greywage                                                                 **
## traffic                                                                  ***
## coach                                                                    *
## way                                                                      .
## extraversion                                                             **
## anxiety_indicator                                                        .
## independ_indicator                                                       *
## selfcontrol_square                                                       *
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model.int1, model.int4)  # significant model
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(stag, event)
##  Model 1: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversi
##  Model 2: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversi
##    loglik  Chisq Df P(>|Chi|)
## 1 -2431.5
## 2 -2442.4 21.891 10   0.01567 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can take it further by noticing that `extraversion:anxiety_indicator:independ_indicator` is not significant in this model (p-value $0.5023 > 0.05$). So, by removing `extraversion:anxiety_indicator:independ_indicator` we can achieve a model with all significant interaction terms. We can compare it to the model without interactions terms and we see that it is significantly different (p-value $0.01567 < 0.05$).

Interpretation and Comments on the Model So, is `model.int4` is our model most fitted to the data? First, `model.int4` has only one interaction term: between all four personality traits, which may imply a vague conclusion. If we consider `model.int3`, there might be interaction between all personality traits, but maybe more so with the three excluding `self control`. Overall, by using `model.int4` we can conclude that a higher score in all four traits, `extraversion`, `anxiety`, `independence`, and `self control`, would result in a higher turnoer rate.

Interaction Terms - among non-personality traits

```
model.int5 <- coxph(formula = Surv(stag, event) ~ strata(profession) +
    greywage + ageentry * way + industry + traffic + coach +
    extraversion + anxiety_indicator + independ_indicator + selfcontrol_square +
    extraversion:anxiety_indicator:independ_indicator:selfcontrol_square,
    data = turnover_new)
anova(model.int5)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(stag, event)
## Terms added sequentially (first to last)
##
##                                                                  loglik
```

```
## NULL                                                                  -2571.6
## greywage                                                              -2564.6
## ageentry                                                              -2501.9
## way                                                                   -2497.8
## industry                                                              -2472.8
## traffic                                                               -2461.0
## coach                                                                 -2456.4
## extraversion                                                          -2451.8
## anxiety_indicator                                                     -2450.1
## independ_indicator                                                    -2448.1
## selfcontrol_square                                                    -2445.9
## ageentry:way                                                          -2437.7
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square -2434.2
##                                                                         Chisq
## NULL
## greywage                                                               13.9826
## ageentry                                                              125.4951
## way                                                                     8.1313
## industry                                                               49.9922
## traffic                                                                23.6452
## coach                                                                   9.1019
## extraversion                                                            9.3533
## anxiety_indicator                                                       3.2940
## independ_indicator                                                      4.0933
## selfcontrol_square                                                      4.4158
## ageentry:way                                                           16.2822
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square    7.1065
##                                                                       Df
## NULL
## greywage                                                               1
## ageentry                                                               1
## way                                                                    2
## industry                                                              15
## traffic                                                                7
## coach                                                                  2
## extraversion                                                           1
## anxiety_indicator                                                      1
## independ_indicator                                                     1
## selfcontrol_square                                                     1
## ageentry:way                                                           2
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square   1
##                                                                       Pr(>|Chi|)
## NULL
## greywage                                                              0.0001845
## ageentry                                                              < 2.2e-16
## way                                                                   0.0171522
## industry                                                              1.208e-05
## traffic                                                               0.0013150
## coach                                                                 0.0105571
## extraversion                                                          0.0022258
## anxiety_indicator                                                     0.0695315
## independ_indicator                                                    0.0430532
## selfcontrol_square                                                    0.0356077
## ageentry:way                                                          0.0002913
```

```
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square  0.0076806
##
## NULL
## greywage                                                              ***
## ageentry                                                              ***
## way                                                                   *
## industry                                                              ***
## traffic                                                               **
## coach                                                                 *
## extraversion                                                          **
## anxiety_indicator                                                     .
## independ_indicator                                                    *
## selfcontrol_square                                                    *
## ageentry:way                                                          ***
## extraversion:anxiety_indicator:independ_indicator:selfcontrol_square **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model.int4, model.int5)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(stag, event)
##  Model 1: ~ strata(profession) + ageentry + industry + greywage + traffic + coach + way + extraversio
##  Model 2: ~ strata(profession) + greywage + ageentry * way + industry + traffic + coach + extraversio
##    loglik Chisq Df P(>|Chi|)
## 1 -2442.4
## 2 -2434.2 16.53  2 0.0002574 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find a significant interaction of `ageentry:way`. Since `way` is not as significant as `ageentry`, I infer that `ageentry` is influencing the effect `way` has, not vice versa. There is a significant difference between models, so we could proceed with including the term `ageentry:way`.

Is this model the best fit? It is hard to tell. Considering there are a lot of parameters now, it may be making the model too complicated. We can probably remove the non-personality interaction terms, but being aware of these interactions is interesting.

Final Model

Thus, we can consider `coxph(formula = Surv(stag, event) ~ strata(profession) + greywage + ageentry*way + industry + traffic+coach + extraversion+anxiety_indicator + independ_indicator + selfcontrol_square + extraversion:anxiety_indicator:independ_indicator:selfcontrol_square, data = turnover_new)` as our final most significant model.