OLIVIA HYUNYOUNG SHIN

Analysis of Employee Turnover Intention with Regression Models

What are the Causes of Employee Turnover? - with focus on compensation and reward, work intensity, and opportunity for growth.

< Introduction >

Half of the US workers consider a job change and watch a job market or look for a job actively according to Gallup Workforce Panel Study. This is a major concern for employers because employee's turnover can cause tremendous amount of cost. Not only do they have to pay direct exit costs, such as severance pay or unemployment taxes, but they also should deal with recruitment costs, training cost, and possibly decreased productivity and morale as well. Thus, it is important to find out what really drives job-hopping among employees.

There have been several reasons for employee turnover. People may have turnover intention due to high workload. They might also look for higher salary or for better working culture and environment. It could also be for expanding their knowledge or developing their skills. This paper aims to figure out the causes of making people find a new job. Hence, turnover intention is a dependent variable. Some possible factors are chosen as independent variables, which is related to compensation and reward, work intensity, and opportunity for growth.

Scholars have shown that compensation and reward, work intensity and opportunity for growth are related to turnover intention. Rewards and salary were negatively related to employee turnover intention (Cao et al., 2013). Heavy workload is one of the most common reasons for turnover, as it has been proved to have notably positive relationship with turnover intention (Xiaoming, 2014). Opportunity to growth can also be one of the main factors for lowering employee turnover rate. Nouri and Parker (2013) stated that career growth opportunities led to lower turnover intentions.

< Hypotheses >

1. The more compensation and rewards employees have, the less turnover intention would be.
2. The higher workload the workers have, the higher turnover intention would be.
3. The more opportunity for growth employees have, the less turnover intention would be.

OLIVIA HYUNYOUNG SHIN

< Dependent Variable >

The dataset used in this project is from General Social Survey (GSS).

Here is the dependent variable, the question associated with it, and how I recoded it.

■ Turnover Intention

Turnover: *trynewjb* has been recoded into *turnover*, with the reversed order.

Taking everything into consideration, how likely is it you will make a genuine effort to find a new job with another employer within the next year?

1 = Not at all likely, 2 = Somewhat likely, 3 = Very likely

< Independent Variables >

Here are the dependent variables, questions associated with each of them, and how I recoded them.

■ Compensation and reward

Fringeok

My fringe benefits are good.

1) Very true 2) Somewhat true 3) Not too true 4) Not at all true

Rincblls

Do you feel that the income from your job alone is enough to meet your family's usual monthly expenses and bills?

0) Yes 1) No

■ Work Intensity

moredays

How many days per month do you work extra hours beyond your usual schedule?

0~30 days per month

Wkvsfam has been recoded into the reversed order.

How often do the demands of your job interfere with your family life?

1) Never 2) Rarely 3) Sometimes 4) Often

■ Opportunity for Growth

myskills

My job lets me use my skills and abilities.

1) Strongly agree 2) Agree 3) Disagree 4) Strongly disagree

opdevel

I have an opportunity to develop my own special abilities.

OLIVIA HYUNYOUNG SHIN

1) Very true 2) Somewhat true 3) Not too true 4) Not at all true

< Descriptive Statistics >

Table 1

| | vars | n | mean | sd | min | max | range | se | IQR |
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
wrkstat	1	1441	1.20	0.46	1	3	2	0.01	0
turnover	2	1441	1.57	0.78	1	3	2	0.02	1
fringeok	3	1441	2.10	1.09	1	4	3	0.03	2
rincblls	4	1441	0.51	0.50	0	1	1	0.01	1
moredays	5	1441	6.14	7.86	0	30	30	0.21	10
wkvsfam	6	1441	2.28	0.98	1	4	3	0.03	2
myskills	7	1441	1.73	0.70	1	4	3	0.02	1
opdevel	8	1441	1.91	0.89	1	4	3	0.02	1

The variable turnover shows that respondents have turnover intention with the mean 1.57, which lies between "Not at all likely" and "Somewhat likely". For independent variables, I recoded some of them to have reverse order so the larger value they have, the more I assume that they are likely to seek a new job. fringeok and rincblls indicate how satisfied the respondents are with their current job from an economic point of view. On average, they marked 2.10 out of 4 on their fringe benefit, which is about benefits in addition to the employee's wage, like health insurance. The mean of Rincblls is 0.51. This indicates that about half of the respondents think that the income from their job alone is enough to meet their family's usual monthly expenses and bills. Moredays shows work intensity with mean score at 6.14 out of 30. Wkvsfam is another indicator for work intensity to see whether their job interferes with their family life and has 2.28 mean score out of 4. For the opportunity to grow at work, myskills and opdevel has 1.73 and 1.91 mean out of 4 respectively. Fringeok and wkvsfam have high standard deviations, which means they are spread out from the mean. For instance, I assume that wkvsfam has high standard deviation because it may depend on whether he/she has their own family.

OLIVIA HYUNYOUNG SHIN

< Initial Model: Multiple Linear Regression >

Table 2

```
Call:
lm(formula = turnover ~ fringeok + rincblls + wkvsfam + moredays +
    myskills + opdevel, data = d2, subset = wrkstat == 1 & 2 &
    3)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2346 -0.5339 -0.3092  0.4633  1.8641

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.930925   0.089262  10.429  < 2e-16 ***
fringeok     0.133746   0.021596   6.193 8.12e-10 ***
rincblls     0.082124   0.043522   1.887   0.0594 .
wkvsfam     -0.024299   0.022488  -1.081   0.2801
moredays    -0.001500   0.002732  -0.549   0.5830
myskills     0.142822   0.034738   4.111 4.20e-05 ***
opdevel      0.070603   0.027658   2.553   0.0108 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7292 on 1183 degrees of freedom
Multiple R-squared:  0.08719,   Adjusted R-squared:  0.08256
F-statistic: 18.83 on 6 and 1183 DF,  p-value: < 2.2e-16
```

Table 2 presents the result of multiple linear regression model. As I recoded some independent variables and the dependent variable, I expected all of the independent variables would show positive direction with "turnover". However, some of them have negative relationship with the dependent variable.

First of all, all the variables for compensation and reward, "fringeok" and "rincblls" have positive relationship with "turnover". Controlling for other variables, as moving up one "fringeok" category (worse fringe benefits), the person will score 0.134 higher points on the turnover intention scale, on average. Net of other variables, the respondents who think their income is not enough, compared to those who think their income is enough, will score 0.082 higher points on the turnover intention scale. While they all have statistically significant results, it is interesting to see that "fringeok" has a little bit stronger relationship with turnover than "rincblls". This might imply that fringe benefits are better incentives than salary-based compensation alone.

Compared to compensation and reward related factors, work intensity seems negatively related to turnover intention, which is against what I initially expected. I thought the higher level of workload would cause higher turnover intention. However, in the initial model, as one day increase in "moredays", the

respondents will score 0.0017 less points on "turnover" scale on average, net of other variables, though it is not a statistically significant result.

For variables related to Opportunity for growth, whether it is possible to use their skills and abilities at work, "myskills" seems more positively correlated to turnover than whether the respondents can develop their own special abilities, "opdevel". Given the reversed order of the variables, this means that the less they can use their skills and abilities and the less they have the chance to develop their own special abilities, the more likely they would make a genuine effort to find a new job with another employer within the next year.

Although this initial model has the result that I expected in some of the variables, it has some limitations. Among variables, some of them are under similar type of question, which represents they might be highly correlated. For instance, "moredays", the number of days per month one works extra hours, and "wkvsfam", how often the demands of one's job interfere with their family life, may tap the same underlying concept, and thus are highly correlated. In addition, this model did not address other possible variables that could have an impact on both independent and dependent variables. For example, respondents' marital status might have a relationship with turnover intention, whether he/she feels that the income is enough to meet their family's expenses and bills or not ("rincblls"), or how often the demands of job interfere with his/her family life ("wkvsfam"). Also, "income" might have an impact on the respondents' turnover intention as well as "rincblls". These problems could be improved by creating scales for some independent variables and by adding control variables, that might drive changes both in dependent and independent variables.

< Model improvement 1: Multicollinearity Test >

Although the initial model generated meaningful results with overall decent statistical significance, multicollinearity might have occurred due to some variables under similar questions. Therefore, I would like to check it with two methods, Variance Inflation Factor (VIF) and Cronbach's Alpha score.
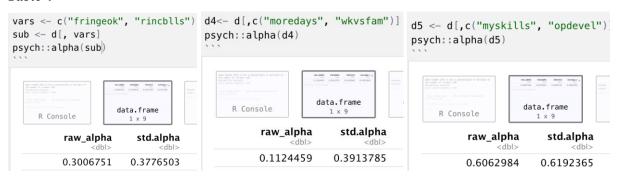
Table 3

```
car::vif(lm00)
```

```
 fringeok rincblls  wkvsfam moredays myskills  opdevel
 1.092761 1.055858 1.067547 1.083087 1.237910 1.295865
```

First, I have checked it with the variance inflation factor (VIF) scores. It measures the extend of correlation between one predictor and the others in a model to diagnose multicollinearity. Higher values signify less accurate contribution of the predictor to a model. According to the VIF scores, all variables seemed reasonable to be included. Because the VIF values are all between 1.08 and 1.37, I assumed that the multicollinearity issues are not present in my model.

Table 4

```
vars <- c("fringeok", "rincblls")       d4<- d[,c("moredays", "wkvsfam")]      d5 <- d[,c("myskills", "opdevel")]
sub <- d[, vars]                         psych::alpha(d4)                       psych::alpha(d5)
psych::alpha(sub)
```

raw_alpha <dbl>	std.alpha <dbl>	raw_alpha <dbl>	std.alpha <dbl>	raw_alpha <dbl>	std.alpha <dbl>
0.3006751	0.3776503	0.1124459	0.3913785	0.6062984	0.6192365

Cronbach's Alpha scores for the independent variables in Table 4 measures internal consistency reliability. Its score goes up if the two variables measure the same underlying concept, but goes down if they don't. The correlation between "fringeok" and "rincblls" is 0.3. "moredays" and "wkvsfam" has 0.39 standardized alpha score. Although "myskills" and "opdevel" seems to be correlated with about 0.6 alpha score, we need 0.7 or higher to conclude high enough internal consistency. Since there were not significantly high correlations between those variables, creating a new scale for my independent variables does not seem necessary.

Table 5

```
Call:
lm(formula = turnover ~ heavywork + reward + growth, data = d,
    subset = wrkstat == 1 & 2 & 3)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1777 -0.5181 -0.3185  0.4775  1.7983

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.56827    0.02138  73.354  < 2e-16 ***
heavywork    0.01038    0.03405   0.305     0.76
reward       0.17877    0.02853   6.267 5.16e-10 ***
growth       0.17487    0.02629   6.651 4.43e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7316 on 1186 degrees of freedom
Multiple R-squared:  0.0789,    Adjusted R-squared:  0.07657
F-statistic: 33.87 on 3 and 1186 DF,  p-value: < 2.2e-16
```

When I tried to create a scale and made new linear regression model, I was able to notice a change on a coefficient. "heavywork" is a new scale for "moredays and wkvsfam", "reward" is a new scale for "fringeok" and "rincblls", and "growth" is for "myskills" and "opdevel". "heavywork" for measuring the work intensity became positive (+0.01038), whereas "wkvsfam" (-0.024) and "moredays"(-0.015) were negative in the initial model. Despite the fact that the model with new scales generated all positive directions, which I originally expected, lower adjusted R-squared score and insufficient evidence for creating new scales led me to choose the initial model.

< Model improvement 2: Adding control variables >

■ Control variables

- *Race* has been recoded into 1 = black and other, 0= white

What race do you consider yourself?

1) White 2) Black 3) Other

- *Rincome:* 13) refused and 98) Don't know have been removed

In which of these groups did your earnings from (STATED OCCUPATION) for last year fall? That is, before taxes or other deductions.

1) Under $1,000 2) $1,000 to 2,999 3) $3,000 to 3,999……11) $20,000 to 24,999, 12) $25,000 or over, 13) Refused 98) Don't know)

- *Age:* Respondent's age

- *Wrkstat:* I subsetted 1,2,3 to see employees only.

Last week were you working full-time, part-time, going to school, keeping house, or what?

1) Working full-time 2) Working part-time 3) With a job, but not at work because of temporary illness, vacation, strike 4) Unemployed, laid off, looking for work 5) Retired 6) In school 7) Keeping house 8) Other 9) No answer

- *Marital* has been recoded into 1 = married, 0= all the others

Are you currently -- married, widowed, divorced, separated, or have you never been married?

1) Married 2) Widowed 3) Divorced 4) Separated 5) Never Married 9) No answer

<Descriptive Statistics for Control Variables>

Table 6

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	min <dbl>	max <dbl>	range <dbl>	se <dbl>	IQR <dbl>
income	9	1441	11.41	1.66	1	12	11	0.04	0
age	10	1441	41.66	12.52	18	88	70	0.33	19
race	11	1441	0.15	0.35	0	1	1	0.01	0
marital	12	1441	0.49	0.50	0	1	1	0.01	1

For control variables, income has very high mean, 11.41, considering its max value is 12. This is because more than half of the respondents selected 12, $25,000 or over as their last year income. Race has low mean value because I recoded 1 as Black or Other and they are about 27% of the total respondents

according to GSS, whereas 0 is White, and they are about 73%. Married respondents were about 48% and so its mean value is 0.49.

Table 7

```
Call:
lm(formula = turnover ~ fringeok + rincblls + moredays + wkvsfam +
    myskills + opdevel + race + income + age + marital, data = d2,
    subset = wrkstat == 1 & 2 & 3)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4732 -0.4956 -0.2461  0.4352  1.9841

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.9536126  0.1991367   9.810  < 2e-16 ***
fringeok     0.1147113  0.0212913   5.388 8.61e-08 ***
rincblls     0.0819820  0.0431831   1.898  0.05788 .
moredays    -0.0004894  0.0026605  -0.184  0.85409
wkvsfam     -0.0156810  0.0221136  -0.709  0.47840
myskills     0.1102848  0.0340638   3.238  0.00124 **
opdevel      0.0850939  0.0270201   3.149  0.00168 **
race         0.1296691  0.0589263   2.201  0.02796 *
income      -0.0446990  0.0144875  -3.085  0.00208 **
age         -0.0113894  0.0018001  -6.327 3.54e-10 ***
marital     -0.0364224  0.0439665  -0.828  0.40760
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7093 on 1179 degrees of freedom
Multiple R-squared:  0.1392,    Adjusted R-squared:  0.1319
F-statistic: 19.07 on 10 and 1179 DF,  p-value: < 2.2e-16
```

Table 7 included control variables; race, income, age, and marital status. It allows me to identify the relationship between the demographic variables and the dependent variable, "turnover". Most of them, race, income, and age have generated statistically significant results. Holding other variables fixed, if the respondents are non-White, compared to the White, will score 0.129 higher points on the turnover intention scale. Net of other variables, getting one year older will lead to 0.0113 lower points on the turnover intention scale. Although statistically not significant, compared to those who are not married, the respondents who are married will score 0.036 less points on the turnover intention scale. This might indicate that if he/she is married, they are less likely to try to change their job, possibly because they prefer stability. Being in one higher category in income will lead to about 0.045 lower points on the turnover intention scale, indicating that higher salary might result in lower turnover rate.

It is interesting to see that among all independent variables, including control variables, "fringeok" and "myskills" are the two most highly positively correlated variables with "turnover". This implies that reward and the chance to use their own skills and ability are really important factors that impact employee

turnover decision. In addition, the coefficient of "fringeok" and "rincblls" decreased compared to the initial model, which might be because the part of associations between "fringeok" and "turnover", and between "rincblls" and "turnover" are related to exogenous variable, such as "income".

Moreover, adjusted R-squared went up to 0.1319 from 0.08256. As the adjusted R-squared is a modified version of R-squared, which increases only if the new variables improved the previous model more than would be expected by chance, I assume that the control variables improved my initial model.

< Model improvement 3: Interaction >

Table 8

```
Call:
lm(formula = turnover ~ fringeok + rincblls + moredays + wkvsfam +
    myskills + opdevel + race + income + age + marital + marital:wkvsfam,
    data = d2, subset = wrkstat == 1 & 2 & 3)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4735 -0.4975 -0.2479  0.4353  1.9895

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.9600206  0.2028824   9.661  < 2e-16 ***
fringeok       0.1147645  0.0213025   5.387 8.63e-08 ***
rincblls       0.0817620  0.0432210   1.892  0.05877 .
moredays      -0.0005131  0.0026654  -0.193  0.84738
wkvsfam       -0.0188612  0.0291930  -0.646  0.51835
myskills       0.1102105  0.0340808   3.234  0.00126 **
opdevel        0.0851528  0.0270336   3.150  0.00167 **
race           0.1296424  0.0589508   2.199  0.02806 *
income        -0.0446631  0.0144951  -3.081  0.00211 **
age           -0.0113761  0.0018026  -6.311 3.91e-10 ***
marital       -0.0532374  0.1098994  -0.484  0.62818
wkvsfam:marital 0.0071534  0.0428456   0.167  0.86743
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7096 on 1178 degrees of freedom
Multiple R-squared:  0.1392,    Adjusted R-squared:  0.1312
F-statistic: 17.32 on 11 and 1178 DF,  p-value: < 2.2e-16
```

This model shows the result of the model where the interaction term is included. The coefficient of "wkvsfam:marital" is 0.007, though it is not statistically significant. This might demonstrate that the impact of "wkvsfam" on the turnover intention can be different depending on the respondent's marital status. I previously expected that if he/she is married, he/she is more likely to dislike their family life being interfered with heavy workload at work, and thus more likely to have higher turnover intention under heavy workload. Looking into this model, all else unchanged, if the respondent is married, one unit increases in "wkvsfam" will decrease 0.0118 (0.007 – 0.0188) points on turnover intention scale, on average. On the other hand, those who are not married, will decrease 0.0188 points as one unit increases in "wkvsfam", on average, ceteris paribus. Although those who are married and not married both have

negative relationship with turnover intention, I noticed that if they are married and their family life is interfered with workload, they are slightly more likely to try to look for a new job.

However, in terms of adjusted R-squared, this new model has not been improved from the model without the interaction term. Therefore, when choosing between the two, Table 7 with 0.1319 adjusted R-squared and Table 8 with 0.1312 adjusted R-squared, I took the former because of the stronger explanatory power.

< Model improvement 4: Ordered Logistic Model – Final Model >

Table 9

```
formula:
as.factor(turnover) ~ fringeok + rincblls + moredays + wkvsfam + myskills +
opdevel + race + income + age + marital
data:    d2


Coefficients:
          Estimate Std. Error z value Pr(>|z|)
fringeok  0.312567   0.053568   5.835 5.38e-09 ***
rincblls  0.368705   0.117143   3.147 0.001647 **
moredays -0.004154   0.007539  -0.551 0.581645
wkvsfam   0.021667   0.059190   0.366 0.714316
myskills  0.315218   0.086826   3.630 0.000283 ***
opdevel   0.278241   0.069619   3.997 6.43e-05 ***
race      0.308268   0.153824   2.004 0.045066 *
income   -0.082135   0.033141  -2.478 0.013198 *
age      -0.038471   0.004726  -8.141 3.92e-16 ***
marital  -0.176630   0.119403  -1.479 0.139067
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I have tried to improve my model by generating ordered logistic model with the same variables. The interaction term was excluded because the standard error of ordered logistic model with the interaction term was higher than this model (Table 9). Since the dependent variable consists of discrete categories, if I treat them as continuous variables like in multiple linear regression model, it might lead to large bias. Therefore, I would choose ordered logistic model as my final model because it treats the ordered variables better, and thus better interpretation for the change in categorical dependent variable.

Compared to Table 7, where same variables are included but multiple linear regression is applied, the p-values of each variables are generally decreased, indicating the higher statistical significance of my model. Also, the direction of the wkvsfam's coefficient changed from negative to positive, though not statistically significant. Holding other variables fixed, for one higher category of "wkvsfam" (more often interfere with family life due to demands of the job), the odds that the respondent is in the one higher turnover intention category (among Not at all likely, Somewhat likely, Very likely) are 0.0216 times higher on average. Similar interpretations can be applied to other variables in the model. Unlike in the

multiple linear regression in Table 7, in this model, the coefficient of "rincblls" (about income) is bigger than "fringeok" (about fringe benefit). This shows a higher relationship between enough income and turnover intention than between fringe benefit and turnover intention.

< Conclusion >

The three initial hypotheses are only partially supported with the final model. Unlike my expectation, not enough evidence to assure the positive relationship between high work intensity and turnover intention was present, because of the negative relationship with "moredays" and the weak positive relationship with "wkvsfam". I might be able to assume that even if the respondents have to work extra hours beyond their usual schedule, there is no high tendency for making an effort to find a new job. However, the high p-values for the two variables, "moredays" and "wkvsfam", and the positive direction of the new variable heavywork's coefficient in Table 5 made the assumption uncertain. The other two hypotheses turned out to be convincing with low p-values for all four variables. Compensation and rewards and opportunity for growth were positively related with employee turnover intention with coefficients around 0.3 in the ordered logistic model.

There are other limitations in my model. Although the model included control variables, which are race, income, age, and marital status, other factors could affect the independent and dependent variables. Also, the income variable had more than 50% of the people in the highest category, 12) $25,000. This should be further investigated to completely represent the economic status of all population. In addition, the dependent variable, it is possible that "turnover" does not fully represent all employee's turnover intention. The respondents could have not expressed their turnover intention due to financial crisis. They also might have thought it would be demanding to move to another company, regardless of the compensation and rewards, work intensity, and opportunity for growth at their workplace. Finally, since I only considered those who currently enrolled as an employee (full-time, part-time, with a job but in vacation), my model cannot be generalized to the whole population.

OLIVIA HYUNYOUNG SHIN

<div align="center">**< Reference >**</div>

Cao, Z., Chen, J., & Song, Y. (2013). Does Total Rewards Reduce the Core Employees' Turnover
Intention? *International Journal of Business and Management*, *8*(20). doi: 10.5539/ijbm.v8n20p62

Nouri, H., & Parker, R. J. (2013). Career growth opportunities and employee turnover intentions in public
accounting firms. *The British Accounting Review*, *45*(2), 138–148. doi: 10.1016/j.bar.2013.03.002

Xiaoming, Y. (2014). Effects of Workload on Burnout and Turnover Intention of Medical Staff: A
Study. *Studies On Ethno-Medicine*, *08*(03). doi: 10.31901/24566772.2014/08.03.04