University of Rochester

Academic Year 2022/23 Autumn Semester

GBA462R: Core Statistics for MS Students Using R

Module Convenor: **Hana Choi**

Assignment 5: Employee Retention Group Assignment

Team: 3S
Team Member:
Akib Sarwar
Fan Chen
Jiahao Wu
Sijie Huang
Wenhui Yu

Please read Harvard Case No. 9-602-096 “Store24(A): Managing Employee Retention”.
Please use the companion dataset (Store24Data.xlsx) to answer the following questions:
In the Store24 dataset, you will find the store performance and employee characteristics for the chain of “Store24” convenience stores. In particular, you have information on Sales (fiscal year sales), Profit (fiscal year profit), MTenure (average manager tenure in months, that is, how long they have been with the store), and CTenure (average crew tenure in months). The file also contains data on Store24’s most important site location factors: Comp (number of competitors per 10,000 people in a ½ mile radius), Pop (population within a ½ mile radius), Visiblity (5-point rating on visibility, 5 is highest), PedCount (5-point rating on pedestrian foot traffic, 5 is highest), Hours24 (dummy for being open 24 hours a day), and Res (dummy for being in a residential area rather than industrial). The focus of the problem set is on understanding the role of manager and crew tenure (experience) on store level financial performance. As always, the analysis must be completed in R, with code included.

a. As we are interested in financial performance, we will analyze the role of manager and crew tenure on store level ‘Profit’ (as opposed to ‘Sales’). Run a regression of Profit on the eight explanatory variables listed above (i.e., everything but Sales). What is the interpretation of the coefficient on MTenure? Is it significant at the 1% level?

```
### preparation for dataset
library(readxl)
setwd("/Users/huangsijie/Downloads")
df = read_excel("Store24Data.xlsx")

### a. multiple regression with 8 X
fit = lm(formula = Profit ~ MTenure + CTenure + Pop + Comp + Visibility + PedCount + Res
+ Hours24, data = df)
summary(fit)
### Coefficient of MTenure is 760.993, implying that an additional month of manager
tenure is associated with an increase in profit of 760.993 holding other things constant;
### MTenure is statistically significant at the 1% level because p-value < 0.01
```

b. What is the interpretation of the coefficient on CTenure? Is it significant at the 1% level?

```
### b. Coefficient of CTenure is 944.978, implying that an additional month of manager
tenure is associated with an increase in profit of 944.978 holding other things constant;
### MTenure is not statistically significant at the 1% level because p-value is 0.028400,
p-value > 0.01
```

c. Doucette remarked “"For example, our most recent store manager bonus plan provides a quarterly bonus of 3% of the manager's salary for increasing average crew tenure by 1.38 months during the quarter. It would be great if we could use this data to get some estimate of the actual financial impact of a 1.38-month increase in crew tenure."

Based on the regression analysis in (a), what would your answer be to Doucette’s remark? How would you guide Doucette in determining whether the most recent store manager bonus plan was a successful one?

Hint: You will have to weigh the cost and the benefit to the firm. You might not know the exact cost to the firm, so you might have to make some assumptions about the cost, or consider different cost levels.

```
#quarterly 1.38-month increase in CTenure effect on yearly profit
yr_profit_change<-fit$coefficients["CTenure"]*1.38*4
#maximum quarterly bonus amount
q_bonus<-yr_profit_change/4
#maximum quarterly salary
q_salary<- q_bonus/0.03
```

As the estimated coefficient of CTenure is 944.978, a 1.38-month increase in crew tenure in a quarter will cause an average increase in yearly profit by $944.978*1.38*4=**$5216.279**, given other factors constant. To guarantee the manager bonus plan is successful, we should make sure that the cost of giving 3% quarterly bonus of the manager's salary is lower than or equal to the benefit gained from the 1.38-month rise in crew tenure. Hence, the maximum quarterly bonus amount should be **$1304.07** and the maximum manager's quarterly salary should be **$43468.99.**

d. What is the interpretation (in words) of the coefficient estimate on Hours24?

Having other things constant, if a store open 24 hours a day, the profit of a fiscal year before overhead allocation, rent and depreciation will averagely increase by $63233.307 rather than $0 profit gained when not opening 24 hours a day.

e. Based on the regression analysis in (a), construct a 95% confidence interval for the impact of population (Pop) on profit. What, in words, is the interpretation of this interval?

```
> confint.default(reg, level=0.95)["Pop",]
    2.5 %    97.5 %
0.7926957 6.5405168
```

The true effect of Pop on profit is between 0.79 and 6.54 with 95% probability.

f. What is the role of the site location factors in this analysis? What is the expected relationship between performance and each of the site location factors? What happens when you drop all the site location factors from the model? What does this tell you? Comparing the model in (a) and the model without any site location factors, why should (or shouldn't) the site location factors be included in the analysis?

```
> reg = lm(Store24Data$Profit~CTenure + MTenure + Comp + Pop + Visibility + PedCount + Hours24 + Res, data = Store24Data)
> summary(reg)

Call:
lm(formula = Store24Data$Profit ~ CTenure + MTenure + Comp +
    Pop + Visibility + PedCount + Hours24 + Res, data = Store24Data)

Residuals:
    Min      1Q  Median      3Q     Max
-105789  -35946   -7069   33780  112390

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   7610.041  66821.994   0.114 0.909674
CTenure        944.978    421.687   2.241 0.028400 *
MTenure        760.993    127.086   5.988 9.72e-08 ***
Comp         -25286.887   5491.937  -4.604 1.94e-05 ***
Pop              3.667      1.466   2.501 0.014890 *
Visibility   12625.447   9087.620   1.389 0.169411
PedCount     34087.359   9073.196   3.757 0.000366 ***
Hours24      63233.307  19641.114   3.219 0.001994 **
Res          91584.675  39231.283   2.334 0.022623 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56970 on 66 degrees of freedom
Multiple R-squared:  0.6379,    Adjusted R-squared:  0.594
F-statistic: 14.53 on 8 and 66 DF,  p-value: 5.382e-12
```

```
> reg2 = lm(Store24Data$Profit~MTenure + CTenure, data = Store24Data)
> summary(reg2)

Call:
lm(formula = Store24Data$Profit ~ MTenure + CTenure, data = Store24Data)

Residuals:
    Min      1Q  Median      3Q     Max
-165442  -49679   -7474   48733  194710

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 236950.7    13124.7  18.054  < 2e-16 ***
MTenure        619.8      166.7   3.718 0.000394 ***
CTenure        810.1      543.2   1.491 0.140238
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 80210 on 72 degrees of freedom
Multiple R-squared:  0.2168,    Adjusted R-squared:  0.195
F-statistic: 9.965 on 2 and 72 DF,  p-value: 0.0001512
```

When we want to study the relationship between the variables, we should control the other factors the same, we use site location factors here to avoid omitted variable bias. Performance would have the negative relationship with the factor of competitors, while it would have the positive relationship with the other factors: Pop, Visible, PedCount, Hours24 and Res. When drop all the site location factors from the model, Adj-R2 dropped from 0.594 to 0.195, which indicates model fit is worse. Omitting location -based variables causes omitted variable bias by comparing point estimates of MTenure and CTenure change, also lowers estimate precision according to the change in CTenure's standard error, therefore, we should include the site location factors.

g. The Harvard Case reads "Jenkins understood that Hart was essentially saying that the *relationship between tenure and financial performance might vary with the level of tenure*. She recalled that this meant there could be a nonlinear relationship between tenure and financial performance"

Why do you think Jenkins suspects a nonlinear relationship between tenure and financial performance? What is the economic logic (in words)?

Can you incorporate this nonlinear relationship into the model? In other words, can you construct one single model in which (i) the relationship between financial performance and manager tenure varies with the level of manager tenure and also (ii) the relationship between financial performance and crew tenure varies with the level of crew tenure? (so consider building one single model that incorporates both criteria). Based on your answer from (f), decide whether you also want to include the site location factors in this model. What do you find?

Hint: Check the entries for manager tenure. Do you find entries with zero values? Then think about which regression model you should use.

Because firstly when we use the linear model, the consequence of R square is not that good, which means the linear model cannot describe the relationship between tenure and profit well. And also, from the perspective of economic, Limbach et al. (2015) hypothesized, and demonstrated an inverted U-shape relationship between CEO tenure and firm performance. I believe the relationship between tenure of manager and client and firm performance also follows that relationship. This is because in the first period of time of tenure, with the tenure increasing, the experience of workers accumulates. In this case, the financial performance becomes better because people are more efficient in the

management and execution part. As the tenure reaches some point, it's hard for the workers to have a better working experience continuously, which means the managers and the clients have formed some routine of work and follow one's routine even though the tenure increases. And this will lead to no changes or extremely gentle changes in terms of financial performance. When it comes to the workers with even longer tenure than that value, with the tenure increasing, people tend to avoid the changes in working and fail to keep pace with the cutting-edge technology during the work. As a result, the financial performance will even be worse in this period. Therefore, this is why the tenure and financial performance in an inverted U-shape relationship.

In this condition, considering the inverted U-shape, I use the quadratic model of square component of tenure. And I added the location factors so that both the R square and adj R square increases (R square from 0.237 to 0.7045, adj R square from 0.1934 to 0.6583)

```
fit5=lm(Profit~MTenure+I(MTenure**2)+CTenure+I(CTenure**2)+Pop+Comp+Visibility+PedCount
+Res+Hours24,data=data2)
summary(fit5)
################################## RESULT
Residuals:
    Min     1Q Median     3Q    Max
-93706 -29284  -7132  35856 124337

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -24181.491  62263.640  -0.388  0.69903
MTenure        1732.548    312.207   5.549 5.88e-07 ***
I(MTenure^2)     -4.830      1.443  -3.348  0.00137 **
CTenure        2460.359    974.711   2.524  0.01409 *
I(CTenure^2)    -16.901     10.376  -1.629  0.10826
Pop               3.475      1.397   2.487  0.01550 *
Comp         -25455.638   5136.143  -4.956 5.56e-06 ***
Visibility    19761.671   8750.212   2.258  0.02734 *
PedCount      39140.913   8486.859   4.612 1.96e-05 ***
Res           62707.989  38571.379   1.626  0.10891
Hours24       56375.901  18193.912   3.099  0.00289 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 52260 on 64 degrees of freedom
Multiple R-squared:  0.7045, Adjusted R-squared:  0.6583
F-statistic: 15.26 on 10 and 64 DF,  p-value: 1.744e-13
fit6 = lm(Profit~MTenure+I(MTenure**2)+CTenure+I(CTenure**2), data = data2)
summary(fit6)
################################## RESULT
Residuals:
    Min      1Q  Median      3Q     Max
-180663  -45858   -4786   48345  182214

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  221974.581  18865.982  11.766   <2e-16 ***
MTenure        1176.432    457.871   2.569   0.0123 *
I(MTenure^2)     -2.688      2.058  -1.306   0.1958
CTenure        1279.223   1384.874   0.924   0.3588
I(CTenure^2)     -4.843     13.996  -0.346   0.7304
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 80290 on 70 degrees of freedom
```

```
Multiple R-squared:  0.237,  Adjusted R-squared:  0.1934
F-statistic: 5.436 on 4 and 70 DF,  p-value: 0.000719
```

h. Can you use the results (estimates) from (g) to find the point at which an extra month of manager tenure leads to lower profit? If yes, please do so. If not, explain why you cannot. Hint: Write down the profit function with respect to manager tenure. You can treat other factors as constants or ignore them from the profit function. Then optimize this profit function with respect to manager tenure (remember PS3 #2(b)?)

Since we only focus on the change of manager tenure(MTenure), we consider the other parameters as a constant. Therefore, we actually optimize the maximum value of the component related to MTenure, which is the function below:

```
fit5=lm(Profit~MTenure+I(MTenure**2)+CTenure+I(CTenure**2)+Pop+Comp+Visibility+PedCount
+Res+Hours24,data=data2)
profit_func3 = function(p){
    return(fit5$coefficients[['MTenure']]*p+p**2*fit5$coefficients[['I(MTenure^2)']])
}## the profit function is Y = 1732.548X - 4.83025X**2
```

Then with this function related only to MTenure, we do the optimization:

```
optimize(profit_func3, c(min(data2$MTenure),max(data2$MTenure)),tol = 0.01, maximum =
T)## 179.3435
```

And we find that when MTenure achieves 179.3435, the profit will start to decrease when the MTenure increases.