hw3

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

df <- read.table("C:/Rstudy/Introduction to Regression Analysis/datasets/salary-1.txt", header = T);  
str(df)

## 'data.frame': 65 obs. of 4 variables:  
## $ salary: num 58.8 34.8 163.7 70 55.5 ...  
## $ Degree: int 3 1 3 3 3 2 1 1 2 3 ...  
## $ Years : num 4.49 2.92 29.54 9.92 0.14 ...  
## $ Num : int 0 0 42 0 0 4 0 0 3 0 ...

a <- factor(df$Degree)  
levels(a) <- c("Bachelor", "Master", "Doctor")  
df$Degree <- a  
str(df)

## 'data.frame': 65 obs. of 4 variables:  
## $ salary: num 58.8 34.8 163.7 70 55.5 ...  
## $ Degree: Factor w/ 3 levels "Bachelor","Master",..: 3 1 3 3 3 2 1 1 2 3 ...  
## $ Years : num 4.49 2.92 29.54 9.92 0.14 ...  
## $ Num : int 0 0 42 0 0 4 0 0 3 0 ...

cor(df[,c(1, 3, 4)])

## salary Years Num  
## salary 1.0000000 0.7416301 0.8302697  
## Years 0.7416301 1.0000000 0.6354267  
## Num 0.8302697 0.6354267 1.0000000

# result 1  
result1 <- lm(salary~Degree + Years + Num, data = df)  
summary(result1) # R^2: 0.8632

##   
## Call:  
## lm(formula = salary ~ Degree + Years + Num, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -29.058 -3.477 -0.915 3.417 36.909   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.4714 2.8691 10.969 5.73e-16 \*\*\*  
## DegreeMaster 10.8120 3.2183 3.360 0.00136 \*\*   
## DegreeDoctor 22.6307 3.4846 6.494 1.81e-08 \*\*\*  
## Years 1.2581 0.2273 5.535 7.23e-07 \*\*\*  
## Num 1.8523 0.2276 8.137 2.86e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.14 on 60 degrees of freedom  
## Multiple R-squared: 0.8633, Adjusted R-squared: 0.8542   
## F-statistic: 94.76 on 4 and 60 DF, p-value: < 2.2e-16

# result 2  
df$log\_sal <- log(df$salary)   
result2 <- lm(log\_sal~Degree+Years+Num, data = df)  
summary(result2) # R^2: 0.8493 -> 지우고 난 후 R^2: 0.8624

##   
## Call:  
## lm(formula = log\_sal ~ Degree + Years + Num, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.34424 -0.09964 -0.02337 0.09274 0.44192   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.507700 0.043677 80.310 < 2e-16 \*\*\*  
## DegreeMaster 0.252689 0.048994 5.158 2.96e-06 \*\*\*  
## DegreeDoctor 0.473102 0.053048 8.918 1.35e-12 \*\*\*  
## Years 0.023880 0.003460 6.901 3.69e-09 \*\*\*  
## Num 0.015499 0.003465 4.472 3.50e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1544 on 60 degrees of freedom  
## Multiple R-squared: 0.8493, Adjusted R-squared: 0.8393   
## F-statistic: 84.55 on 4 and 60 DF, p-value: < 2.2e-16

# multicolinearity  
library(car)

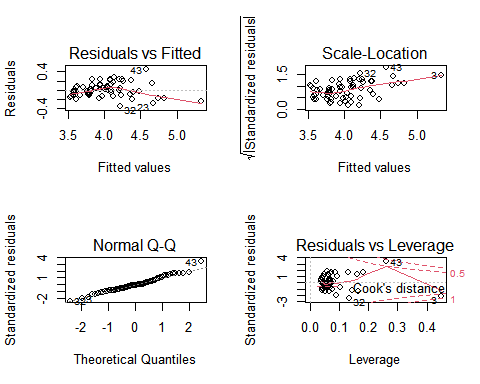
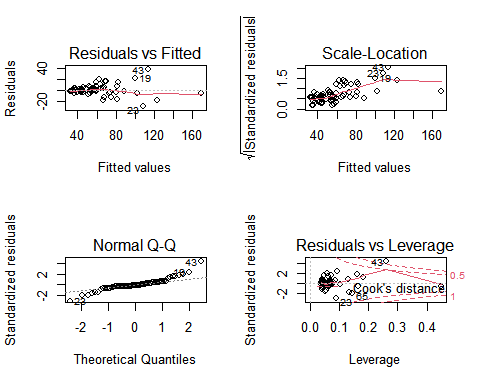
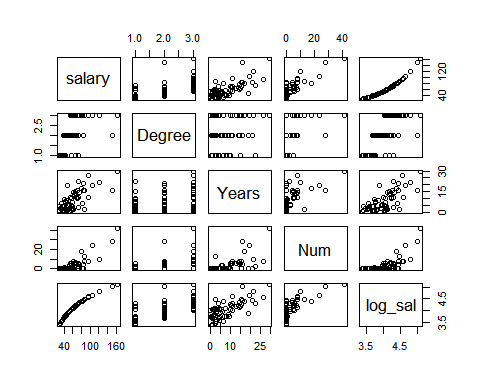
## 필요한 패키지를 로딩중입니다: carData

vif(result2)

## GVIF Df GVIF^(1/(2\*Df))  
## Degree 1.108090 2 1.025991  
## Years 1.685967 1 1.298448  
## Num 1.756278 1 1.325247

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.