9) a) The green part is R code and the screenshot is the R-code output and the black part is my explanation for why I remove the variable.

```
# 9)
# a)
remodel1 <- lm(log(COMP) ~ AGE+EDUCATN+TENURE+EXPER+log(SALES)+log(VAL)+
                    log(PCNTOWN)+log(PROF))
summary(remodel1)
```

```
Call:
lm(formula = log(COMP) ~ AGE + EDUCATN + TENURE + EXPER + log(SALES) +
    log(VAL) + log(PCNTOWN) + log(PROF))

Residuals:
     Min       1Q   Median       3Q      Max
-0.98941 -0.32022 -0.03119  0.23559  1.62794

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.564294   0.959018   5.802 4.44e-07 ***
AGE           0.006135   0.013041   0.470 0.640069
EDUCATN      -0.182187   0.147776  -1.233 0.223394
TENURE        0.001336   0.007346   0.182 0.856431
EXPER        -0.001964   0.010979  -0.179 0.858734
log(SALES)    0.090987   0.088255   1.031 0.307526
log(VAL)      0.442754   0.116585   3.798 0.000397 ***
log(PCNTOWN) -0.388935   0.123770  -3.142 0.002816 **
log(PROF)    -0.164972   0.109882  -1.501 0.139552
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4897 on 50 degrees of freedom
Multiple R-squared:  0.5056,    Adjusted R-squared:  0.4265
F-statistic: 6.392 on 8 and 50 DF,  p-value: 1.067e-05
```

Because p-value of EXPER is the largest, which is 0.858734 and it is larger than 0.15. Thus, we remove EXPER, which is number of years as the firm CEO, and get the following new model.

```
remodel2 <- lm(log(COMP) ~ AGE+EDUCATN+TENURE+log(SALES)+log(VAL)+log(PCNTOWN)
                    +log(PROF))
summary(remodel2)
```

```
Call:
lm(formula = log(COMP) ~ AGE + EDUCATN + TENURE + log(SALES) +
    log(VAL) + log(PCNTOWN) + log(PROF))

Residuals:
    Min      1Q  Median      3Q     Max
-0.9932 -0.3213 -0.0356  0.2333  1.6350

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.601476   0.927301   6.041 1.77e-07 ***
AGE           0.005111   0.011606   0.440  0.66151
EDUCATN      -0.189119   0.141246  -1.339  0.18653
TENURE        0.001039   0.007088   0.147  0.88402
log(SALES)    0.094162   0.085627   1.100  0.27664
log(VAL)      0.441040   0.115083   3.832  0.00035 ***
log(PCNTOWN) -0.392044   0.121376  -3.230  0.00217 **
log(PROF)    -0.166413   0.108541  -1.533  0.13141
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.485 on 51 degrees of freedom
Multiple R-squared:  0.5053,    Adjusted R-squared:  0.4374
F-statistic: 7.442 on 7 and 51 DF,  p-value: 3.727e-06
```

Because p-value of TENURE is the largest, which is 0.88402 and it is larger than 0.15. Thus, we remove TENURE, which is number of years employed by the firm, and get the following new model.

```
remodel3 <- lm(log(COMP) ~ AGE+EDUCATN+log(SALES)+log(VAL)+log(PCNTOWN)
                  +log(PROF))
summary(remodel3)
```

```
Call:
lm(formula = log(COMP) ~ AGE + EDUCATN + log(SALES) + log(VAL) +
    log(PCNTOWN) + log(PROF))

Residuals:
     Min       1Q   Median       3Q      Max
-0.99035 -0.31602 -0.04703  0.21566  1.64058

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.572578   0.897545   6.209    9e-08 ***
AGE           0.005678   0.010841   0.524 0.602709
EDUCATN      -0.193559   0.136657  -1.416 0.162624
log(SALES)    0.095450   0.084370   1.131 0.263111
log(VAL)      0.441011   0.113995   3.869 0.000306 ***
log(PCNTOWN) -0.390445   0.119742  -3.261 0.001964 **
log(PROF)    -0.161889   0.103079  -1.571 0.122356
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4804 on 52 degrees of freedom
Multiple R-squared:  0.5051,    Adjusted R-squared:  0.448
F-statistic: 8.845 on 6 and 52 DF,  p-value: 1.184e-06
```

Because p-value of AGE is the largest, which is 0.602709 and it is larger than 0.15. Thus, we remove AGE,

which is the CEOs age in years, and get the following new model.

```
remodel4 <- lm(log(COMP) ~ EDUCATN+log(SALES)+log(VAL)+log(PCNTOWN)+log(PROF))
summary(remodel4)
```

```
Call:
lm(formula = log(COMP) ~ EDUCATN + log(SALES) + log(VAL) + log(PCNTOWN) +
    log(PROF))

Residuals:
     Min       1Q   Median       3Q      Max
-0.99238 -0.32920  0.00299  0.21677  1.61486

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.93170    0.57511  10.314 2.82e-14 ***
EDUCATN      -0.22244    0.12418  -1.791 0.078950 .
log(SALES)    0.09645    0.08377   1.151 0.254764
log(VAL)      0.44604    0.11281   3.954 0.000229 ***
log(PCNTOWN) -0.39766    0.11813  -3.366 0.001424 **
log(PROF)    -0.16467    0.10223  -1.611 0.113179
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4771 on 53 degrees of freedom
Multiple R-squared:  0.5025,    Adjusted R-squared:  0.4555
F-statistic: 10.71 on 5 and 53 DF,  p-value: 3.815e-07
```

Because p-value of log(SALES) is the largest, which is 0.254764 and it is larger than 0.15. Thus, we remove log(SALES), which is the log transformation of sales revenue in millions of dollar, and get the following new model.

```
remodel5 <- lm(log(COMP) ~ EDUCATN+log(VAL)+log(PCNTOWN)+log(PROF))
summary(remodel5)
```

```
Call:
lm(formula = log(COMP) ~ EDUCATN + log(VAL) + log(PCNTOWN) +
    log(PROF))

Residuals:
     Min       1Q   Median       3Q      Max
-0.98173 -0.31035 -0.06182  0.23900  1.66042

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.4537     0.3549  18.184  < 2e-16 ***
EDUCATN       -0.2311     0.1243  -1.859 0.068451 .
log(VAL)       0.4974     0.1039   4.787 1.36e-05 ***
log(PCNTOWN)  -0.4486     0.1098  -4.084 0.000147 ***
log(PROF)     -0.1476     0.1015  -1.455 0.151415
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4786 on 54 degrees of freedom
Multiple R-squared:   0.49,     Adjusted R-squared:  0.4523
F-statistic: 12.97 on 4 and 54 DF,  p-value: 1.806e-07
```

Because p-value of log(PROF) is the largest, which is 0.151415 and it is larger than 0.15. Thus, we remove log(PROF), which is the log transformation of profits of the firm before taxes in millions of dollars, and get the following new model.

```
remodel6 <- lm(log(COMP) ~ EDUCATN+log(VAL)+log(PCNTOWN))
summary(remodel6)
```

```
Call:
lm(formula = log(COMP) ~ EDUCATN + log(VAL) + log(PCNTOWN))

Residuals:
     Min       1Q   Median       3Q      Max
-0.99808 -0.30991 -0.07039  0.24310  1.63888

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.12699    0.27765  22.068  < 2e-16 ***
EDUCATN      -0.22024    0.12535  -1.757   0.0845 .
log(VAL)      0.37193    0.05856   6.352 4.32e-08 ***
log(PCNTOWN) -0.31235    0.05799  -5.386 1.54e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4834 on 55 degrees of freedom
Multiple R-squared:   0.47,     Adjusted R-squared:  0.4411
F-statistic: 16.26 on 3 and 55 DF,  p-value: 1.094e-07
```

Because p-value of EDUCATN is the largest, which is 0.0845 and it is smaller than 0.15. Thus, we stop since there is no more variables can be removed and we arrive at a reasonable model, whose response variable is the log transformation of CEO compensation in thousands of dollars and explanatory variables include the CEO’s education level, the log transformation of market value of the CEO's stock, and the log transformation of Percentage of firm's market value owned by the CEO.