SMML Class 5 Lab

John Kubale

9/24/2024

Use Income2.csv data with `Income` as the response variable.

```
inc <-read.csv("~/UMD/classes/stat_mod_ML_1_SURV615/class_3/Income2.csv")
```

1. Consider the following three linear regression models (with an intercept). Interpret the coefficient estimates from each model. How do coefficients from models A and B compare to the ones from model C?

A. Income ~ Education - For every one year attained education, income is expected to increase by 6.4. When Education equals 0, we expect the average income to be -41.9

B. Income ~ Seniority - For every one year increase in seniority, income is expected to increase by .3. When Seniority equals 0, we expect the average income to be 39.2.

C. Income ~ Education + Seniority - The coefficient on education from model A is slightly higher than model C, by about .50 income. - The coefficient on seniority from model B is slightly lower than model C, by about .08 income. - The intercept in Model C is lower than both models A and B. - F test, null predictors = to each other and = 0. In one variable, compares predictor with intercept only.

```
m_edu <- lm(Income ~ Education, inc)
summary(m_edu)
```

```
##
## Call:
## lm(formula = Income ~ Education, data = inc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.568  -8.012   1.474   5.754  23.701
##
## Coefficients:
##             Estimate Std. Error t value        Pr(>|t|)
## (Intercept) -41.9166     9.7689  -4.291        0.000192 ***
## Education     6.3872     0.5812  10.990 0.0000000000115 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.93 on 28 degrees of freedom
## Multiple R-squared:  0.8118, Adjusted R-squared:  0.8051
## F-statistic: 120.8 on 1 and 28 DF,  p-value: 0.00000000001151
```

```r
m_sen <- lm(Income ~ Seniority, inc)
summary(m_sen)
```

```
##
## Call:
## lm(formula = Income ~ Seniority, data = inc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.764 -20.232   7.925  20.686  34.622
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 39.15833    8.51594   4.598 0.0000831 ***
## Seniority    0.25129    0.07836   3.207   0.00335 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.51 on 28 degrees of freedom
## Multiple R-squared:  0.2686, Adjusted R-squared:  0.2425
## F-statistic: 10.28 on 1 and 28 DF,  p-value: 0.003347
```

```r
m_both <- lm(Income ~ Education + Seniority, inc)
summary(m_both)
```

```
##
## Call:
## lm(formula = Income ~ Education + Seniority, data = inc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.113 -5.718 -1.095  3.134 17.235
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) -50.08564    5.99878  -8.349 0.00000000585041523 ***
## Education     5.89556    0.35703  16.513 0.00000000000000123 ***
## Seniority     0.17286    0.02442   7.079 0.00000013048839074 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.187 on 27 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9292
## F-statistic: 191.4 on 2 and 27 DF,  p-value: < 0.00000000000000022
```

2. Compute the residual, $\hat{\epsilon}_i$, from a simple linear regression model `Education Seniority`. Regress `Income` on this residual (i.e., fit a model with Income as your outcome and the residuals you calculated as the predictor. What is the estimated slope coefficient? How does this compare to the coefficient estimate from the model #1.C? Given this, what is the meaning of this residual?

- The estimated slope of the residual is 1, which is almost 2 units higher than model C.
- Meaning,

```r
inc <- inc |> mutate(res =  resid(m_both))

m_resid <- lm(Income ~ res, inc)
summary(m_resid)
```

```
##
## Call:
## lm(formula = Income ~ res, data = inc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.832 -23.979   6.222  21.709  39.539
##
## Coefficients:
##             Estimate Std. Error t value         Pr(>|t|)
## (Intercept)  62.7447     4.8510  12.934 0.000000000000249 ***
## res           1.0000     0.7115   1.405             0.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.57 on 28 degrees of freedom
## Multiple R-squared:  0.0659, Adjusted R-squared:  0.03254
## F-statistic: 1.975 on 1 and 28 DF,  p-value: 0.1709
```

NOTE. Useful functions related to `lm`

```r
summary(m_both)
```

```
##
## Call:
## lm(formula = Income ~ Education + Seniority, data = inc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.113 -5.718 -1.095  3.134 17.235
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) -50.08564    5.99878  -8.349 0.00000000585041523 ***
## Education     5.89556    0.35703  16.513 0.00000000000000123 ***
## Seniority     0.17286    0.02442   7.079 0.00000013048839074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.187 on 27 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9292
## F-statistic: 191.4 on 2 and 27 DF,  p-value: < 0.0000000000000022
```

```r
summary.lm(m_both)
```


```
##
## Call:
## lm(formula = Income ~ Education + Seniority, data = inc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.113 -5.718 -1.095  3.134 17.235
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) -50.08564    5.99878  -8.349 0.00000000585041523 ***
## Education     5.89556    0.35703  16.513 0.00000000000000123 ***
## Seniority     0.17286    0.02442   7.079 0.00000013048839074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.187 on 27 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9292
## F-statistic: 191.4 on 2 and 27 DF,  p-value: < 0.0000000000000022
```


```r
library(faraway)
sumary(m_both) # sumary() is from faraway package
```

```
##               Estimate Std. Error t value            Pr(>|t|)
```

```
## (Intercept) -50.085639   5.998779 -8.3493 0.00000000585041523
## Education     5.895556   0.357031 16.5127 0.00000000000000123
## Seniority     0.172855   0.024419  7.0788 0.00000013048839074
##
## n = 30, p = 3, Residual SE = 7.18663, R-Squared = 0.93
```

```r
coef(m_both)
```

```
## (Intercept)   Education   Seniority
## -50.0856388   5.8955560   0.1728555
```

```r
m_both$coefficients
```

```
## (Intercept)   Education   Seniority
## -50.0856388   5.8955560   0.1728555
```

```r
summary(m_both)$coeff
```

```
##                Estimate Std. Error   t value                Pr(>|t|)
## (Intercept) -50.0856388 5.99877911 -8.349305 0.000000005850415225629
## Education     5.8955560 0.35703114 16.512722 0.000000000000001230258
## Seniority     0.1728555 0.02441884  7.078775 0.000000130488390739419
```

```r
vcov(m_both)
```

```
##             (Intercept)    Education     Seniority
## (Intercept) 35.98535084 -1.929595470 -0.0281797521
## Education   -1.92959547  0.127471232 -0.0016958342
## Seniority   -0.02817975 -0.001695834  0.0005962796
```

```r
fitted(m_both)
```

```
##         1         2         3         4         5         6         7         8
##  96.72760  78.28417  38.47238  82.76741  70.87600  62.19073  93.40682  88.92358
##         9        10        11        12        13        14        15        16
##  85.11740  28.42047  39.64737  85.01530  22.08519  64.23443  41.29522  86.19029
##        17        18        19        20        21        22        23        24
##  36.72243  55.65126  78.18208 102.28373  75.06549  59.54700  69.11351  16.91236
##        25        26        27        28        29        30
##  80.63416  23.93724  96.62550  28.52257  26.67052  68.81976
```

```r
m_both$fitted.values
```

```
##         1         2         3         4         5         6         7         8
##  96.72760  78.28417  38.47238  82.76741  70.87600  62.19073  93.40682  88.92358
##         9        10        11        12        13        14        15        16
##  85.11740  28.42047  39.64737  85.01530  22.08519  64.23443  41.29522  86.19029
##        17        18        19        20        21        22        23        24
##  36.72243  55.65126  78.18208 102.28373  75.06549  59.54700  69.11351  16.91236
```

```
##        25        26        27        28        29        30
##  80.63416  23.93724  96.62550  28.52257  26.67052  68.81976
```

```r
predict(m_both)
```

```
##         1         2         3         4         5         6         7         8
##  96.72760  78.28417  38.47238  82.76741  70.87600  62.19073  93.40682  88.92358
##         9        10        11        12        13        14        15        16
##  85.11740  28.42047  39.64737  85.01530  22.08519  64.23443  41.29522  86.19029
##        17        18        19        20        21        22        23        24
##  36.72243  55.65126  78.18208 102.28373  75.06549  59.54700  69.11351  16.91236
##        25        26        27        28        29        30
##  80.63416  23.93724  96.62550  28.52257  26.67052  68.81976
```

```r
residuals(m_both)
```

```
##          1          2          3          4          5          6          7
##  3.1895710 14.2949604 -3.7936533 -4.0646032 -2.8660781  9.3137512 -5.4363491
##          8          9         10         11         12         13         14
## -9.1125513  4.8889307 17.2350590 -7.7335660 11.2676979  5.8973131  2.3673636
##         15         16         17         18         19         20         21
##  0.2367697  2.8104084 -7.9061335  2.0304336 -8.0769810 -3.4497211 -0.3607872
##         22         23         24         25         26         27         28
## -6.0148933  2.9654141  1.6583024 -1.8283771 -2.5486742 -5.8114691 -5.8864053
##         29         30
## -9.0569312  5.7911990
```

```r
m_both$residuals
```

```
##          1          2          3          4          5          6          7
##  3.1895710 14.2949604 -3.7936533 -4.0646032 -2.8660781  9.3137512 -5.4363491
##          8          9         10         11         12         13         14
## -9.1125513  4.8889307 17.2350590 -7.7335660 11.2676979  5.8973131  2.3673636
##         15         16         17         18         19         20         21
##  0.2367697  2.8104084 -7.9061335  2.0304336 -8.0769810 -3.4497211 -0.3607872
##         22         23         24         25         26         27         28
## -6.0148933  2.9654141  1.6583024 -1.8283771 -2.5486742 -5.8114691 -5.8864053
##         29         30
## -9.0569312  5.7911990
```

```r
summary(m_both)$resid
```

```
##          1          2          3          4          5          6          7
##  3.1895710 14.2949604 -3.7936533 -4.0646032 -2.8660781  9.3137512 -5.4363491
##          8          9         10         11         12         13         14
## -9.1125513  4.8889307 17.2350590 -7.7335660 11.2676979  5.8973131  2.3673636
##         15         16         17         18         19         20         21
##  0.2367697  2.8104084 -7.9061335  2.0304336 -8.0769810 -3.4497211 -0.3607872
```

```
##          22          23          24          25          26          27          28
## -6.0148933  2.9654141  1.6583024 -1.8283771 -2.5486742 -5.8114691 -5.8864053
##          29          30
## -9.0569312  5.7911990
```

```r
anova(m_both)
```

```
## Analysis of Variance Table
##
## Response: Income
##           Df  Sum Sq Mean Sq F value                Pr(>F)
## Education  1 17179.3 17179.3 332.625 < 0.00000000000000022 ***
## Seniority  1  2588.0  2588.0  50.109          0.0000001305 ***
## Residuals 27  1394.5    51.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
aov(m_both)
```

```
## Call:
##    aov(formula = m_both)
##
## Terms:
##                 Education Seniority Residuals
## Sum of Squares  17179.303  2588.018  1394.488
## Deg. of Freedom         1         1        27
##
## Residual standard error: 7.186634
## Estimated effects may be unbalanced
```

```r
deviance(m_both) # -2 loglik
```

```
## [1] 1394.488
```

```r
sum((inc$Income-predict(m_both))^2)
```

```
## [1] 1394.488
```

```r
summary(m_both)$fstatistic
```

```
##    value    numdf    dendf
## 191.3669   2.0000  27.0000
```

```r
df.residual(m_both)
```

```
## [1] 27
```

```r
summary(m_both)$df
```

```
## [1]  3 27  3
```

```r
summary(m_both)$r.squared
```

```
## [1] 0.9341035
```

```r
summary(m_both)$adj.r.squared
```

```
## [1] 0.9292223
```

3. Focus on the multiple linear regression model in #1.C. Examine the output.

```r
summary(m_both)
```

```
##
## Call:
## lm(formula = Income ~ Education + Seniority, data = inc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.113 -5.718 -1.095  3.134 17.235
##
## Coefficients:
##              Estimate Std. Error t value           Pr(>|t|)
## (Intercept) -50.08564    5.99878  -8.349 0.00000000585041523 ***
## Education     5.89556    0.35703  16.513 0.00000000000000123 ***
## Seniority     0.17286    0.02442   7.079 0.00000013048839074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.187 on 27 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9292
## F-statistic: 191.4 on 2 and 27 DF,  p-value: < 0.00000000000000022
```

A. Use the code below to calculate t-values and p-values of the slope coefficients "by hand"? What do they allow you to do? Do they match what you get from summary(m_both)?

```r
df_both<-df.residual(m_both)

t_edu<-coef(m_both)[2]/sqrt(vcov(m_both)[2,2])
t_edu
```

```
## Education
##  16.51272
```

```r
p_edu<-2*pt(-abs(t_edu),df_both)
p_edu
```

```
##                Education
```

```
## 0.000000000000001230258
```

```r
t_sen<-coef(m_both)[3]/sqrt(vcov(m_both)[3,3])
t_sen
```

```
## Seniority
##  7.078775
```

```r
p_sen<-2*pt(-abs(t_sen),df_both)
p_sen
```

```
##       Seniority
## 0.0000001304884
```

```r
summary(m_both)
```

```
##
## Call:
## lm(formula = Income ~ Education + Seniority, data = inc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.113 -5.718 -1.095  3.134 17.235
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) -50.08564    5.99878  -8.349 0.00000000585041523 ***
## Education     5.89556    0.35703  16.513 0.00000000000000123 ***
## Seniority     0.17286    0.02442   7.079 0.00000013048839074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.187 on 27 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9292
## F-statistic: 191.4 on 2 and 27 DF,  p-value: < 0.00000000000000022
```

B. Construct 95% confidence interval of β_1.

```r
confint(m_both, 'Education', level=0.95)
```

```
##              2.5 %   97.5 %
## Education 5.162989 6.628123
```

C. How are the residual standard error and its degrees of freedom computed?

- We know $\hat{\sigma} = \sqrt{\dfrac{RSS}{df}}$.

```r
# computes RSS
sqrt(deviance(m_both)/df.residual(m_both))
```

```
## [1] 7.186634
```

```r
summary(m_both)$sigma
```

```
## [1] 7.186634
```

D. How is R^2 calculated? How about adjusted R^2? What do they mean?

- From lecture note p. 54, $R^2 = \frac{SS_{Reg}}{SS_Y} = 1 - \frac{RSS}{SS_Y}$ and $R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-p}$

```r
1 - (sum(inc$res) /
       ( sum( (inc$Income - mean(inc$Income) )^2))
)
```

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

E. What is the F-statistic here? How is this computed? What does it mean?

- From lecture note p. 49, $F = \frac{(SS_Y - RSS)/(p-1)}{RSS/(n-p)}$

```r
anova(m_both)
```

```
## Analysis of Variance Table
##
## Response: Income
##           Df  Sum Sq Mean Sq F value                Pr(>F)
## Education  1 17179.3 17179.3 332.625 < 0.00000000000000022 ***
## Seniority  1  2588.0  2588.0  50.109          0.0000001305 ***
## Residuals 27  1394.5    51.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
F_both<-((sum(anova(m_both)[,2])-anova(m_both)[3,2])/(3-1))/
  (anova(m_both)[3,2]/df.residual(m_both))
F_both
```

```
## [1] 191.3669
```

```r
summary(m_both)
```

```
##
## Call:
## lm(formula = Income ~ Education + Seniority, data = inc)
##
## Residuals:
##    Min      1Q Median      3Q     Max
```

```
## -9.113 -5.718 -1.095  3.134 17.235
##
## Coefficients:
##              Estimate Std. Error t value           Pr(>|t|)
## (Intercept) -50.08564    5.99878  -8.349 0.00000000585041523 ***
## Education     5.89556    0.35703  16.513 0.00000000000000123 ***
## Seniority     0.17286    0.02442   7.079 0.00000013048839074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.187 on 27 degrees of freedom
## Multiple R-squared:  0.9341, Adjusted R-squared:  0.9292
## F-statistic: 191.4 on 2 and 27 DF,  p-value: < 0.00000000000000022
```

F. What does the anova table tell us? -

```r
anova(m_both)
```

```
## Analysis of Variance Table
##
## Response: Income
##           Df  Sum Sq Mean Sq F value                Pr(>F)
## Education  1 17179.3 17179.3 332.625 < 0.00000000000000022 ***
## Seniority  1  2588.0  2588.0  50.109          0.0000001305 ***
## Residuals 27  1394.5    51.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. Test the following for the multiple regression model in #1.3.

A. Are the effects of Education and Seniority the same? I.e., $H_0 : \beta_1 = \beta_2$. What do you think the I() in the formula is doing?

```r
m_both1<-lm(Income~I(Education+Seniority), inc)
summary(m_both1)
```

```
##
## Call:
## lm(formula = Income ~ I(Education + Seniority), data = inc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.024 -20.053   7.381  19.147  34.781
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              32.68707    9.14653   3.574 0.001301 **
```

```
## I(Education + Seniority)  0.27264    0.07407   3.681 0.000982 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.57 on 28 degrees of freedom
## Multiple R-squared:  0.3261, Adjusted R-squared:  0.302
## F-statistic: 13.55 on 1 and 28 DF,  p-value: 0.0009818
```

```r
anova(m_both1)
```

```
## Analysis of Variance Table
##
## Response: Income
##                           Df  Sum Sq Mean Sq F value    Pr(>F)
## I(Education + Seniority)  1  6900.8  6900.8  13.549 0.0009818 ***
## Residuals                 28 14261.0   509.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(m_both1,m_both)
```

```
## Analysis of Variance Table
##
## Model 1: Income ~ I(Education + Seniority)
## Model 2: Income ~ Education + Seniority
##   Res.Df     RSS Df Sum of Sq      F               Pr(>F)
## 1     28 14261.0
## 2     27  1394.5  1     12867 249.12 0.000000000000003728 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

B. Is the slope of Education 6? I.e., $H_0 : \beta_1 = 6$. What do you think the offset() function in the formula is doing?

```r
m_both2<-lm(Income~offset(6*Education)+Seniority, inc)
summary(m_both2)
```

```
##
## Call:
## lm(formula = Income ~ offset(6 * Education) + Seniority, data = inc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.629 -5.657 -1.314  2.810 17.929
##
## Coefficients:
##              Estimate Std. Error t value             Pr(>|t|)
```

```
## (Intercept) -51.66666    2.56024 -20.180 < 0.0000000000000002 ***
## Seniority     0.17147    0.02356   7.278         0.0000000635 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.068 on 28 degrees of freedom
## Multiple R-squared:  0.6542, Adjusted R-squared:  0.6419
## F-statistic: 52.98 on 1 and 28 DF,  p-value: 0.00000006346
```

```r
anova(m_both2)
```

```
## Analysis of Variance Table
##
## Response: Income
##           Df Sum Sq Mean Sq F value        Pr(>F)
## Seniority  1 2646.7 2646.72  52.976 0.00000006346 ***
## Residuals 28 1398.9   49.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(m_both2,m_both)
```

```
## Analysis of Variance Table
##
## Model 1: Income ~ offset(6 * Education) + Seniority
## Model 2: Income ~ Education + Seniority
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 1398.9
## 2     27 1394.5  1    4.4198 0.0856 0.7721
```

5. Is Income ~ Education + Seniority better than Income ~ Education? Use the code below to justify your answer.

- We can examine this with R^2, R^2_{adj}, MSE and General F-test.
- From lecture note p. 51 and 54, General $F = \dfrac{(RSS_{Reduced} - RSS_{Full})/(p-q)}{RSS_{Full}/(n-p)}$ evaluated against F^{p-q}_{n-p}.

```r
summary(m_edu)$r.squared; summary(m_both)$r.squared
```

```
## [1] 0.8118069

## [1] 0.9341035
```

```r
summary(m_edu)$adj.r.squared; summary(m_both)$adj.r.squared
```

```
## [1] 0.8050857

## [1] 0.9292223
```

```r
summary(m_edu)$sigma^2; summary(m_both)$sigma^2
```

```
## [1] 142.2324
```

```
## [1] 51.64771
```

```r
GenF<-((deviance(m_edu)-deviance(m_both))/(3-2))/
  (deviance(m_both)/df.residual(m_both))
GenF
```

```
## [1] 50.10906
```

```r
qf(0.05,3-2,df.residual(m_both), lower.tail = F)
```

```
## [1] 4.210008
```

```r
pf(GenF,3-2,df.residual(m_both), lower.tail = F)
```

```
## [1] 0.0000001304884
```

```r
anova(m_edu, m_both)
```

```
## Analysis of Variance Table
##
## Model 1: Income ~ Education
## Model 2: Income ~ Education + Seniority
##   Res.Df    RSS Df Sum of Sq      F       Pr(>F)
## 1     28 3982.5
## 2     27 1394.5  1      2588 50.109 0.0000001305 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```