NATIONAL RESEARCH UNIVERSITY

Department of Sociology
Laboratory for Comparative Social Research

QUANTITATIVE DATA ANALYSIS

Linear Regression III

Violetta Korsunova

St. Petersburg, 2021

ASSUMPTIONS OF LINEAR REGRESSION

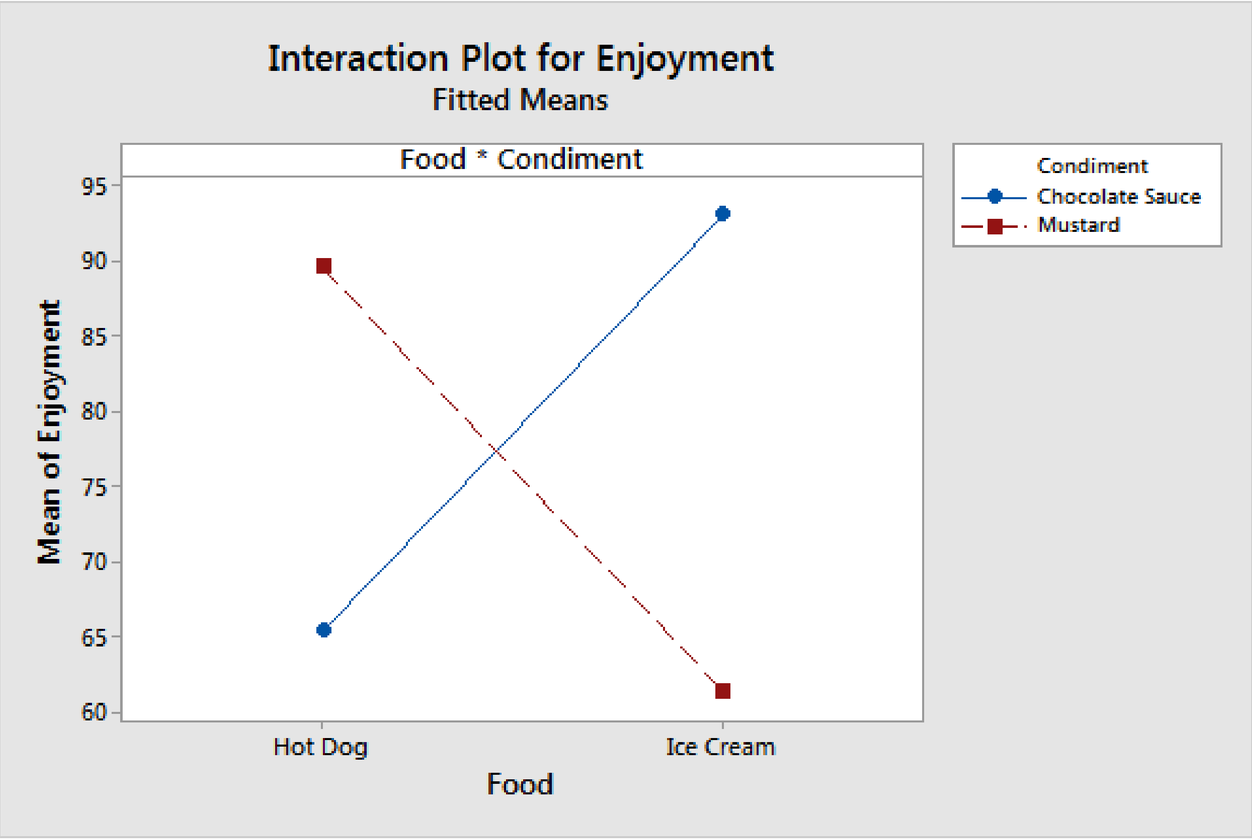
- Independent variables are unrelated
- Effects are unrelated
- The link between the IV and DV is linear
- Dependent variable is normally distributed

PROBLEMS YOU CAN ENCOUNTER

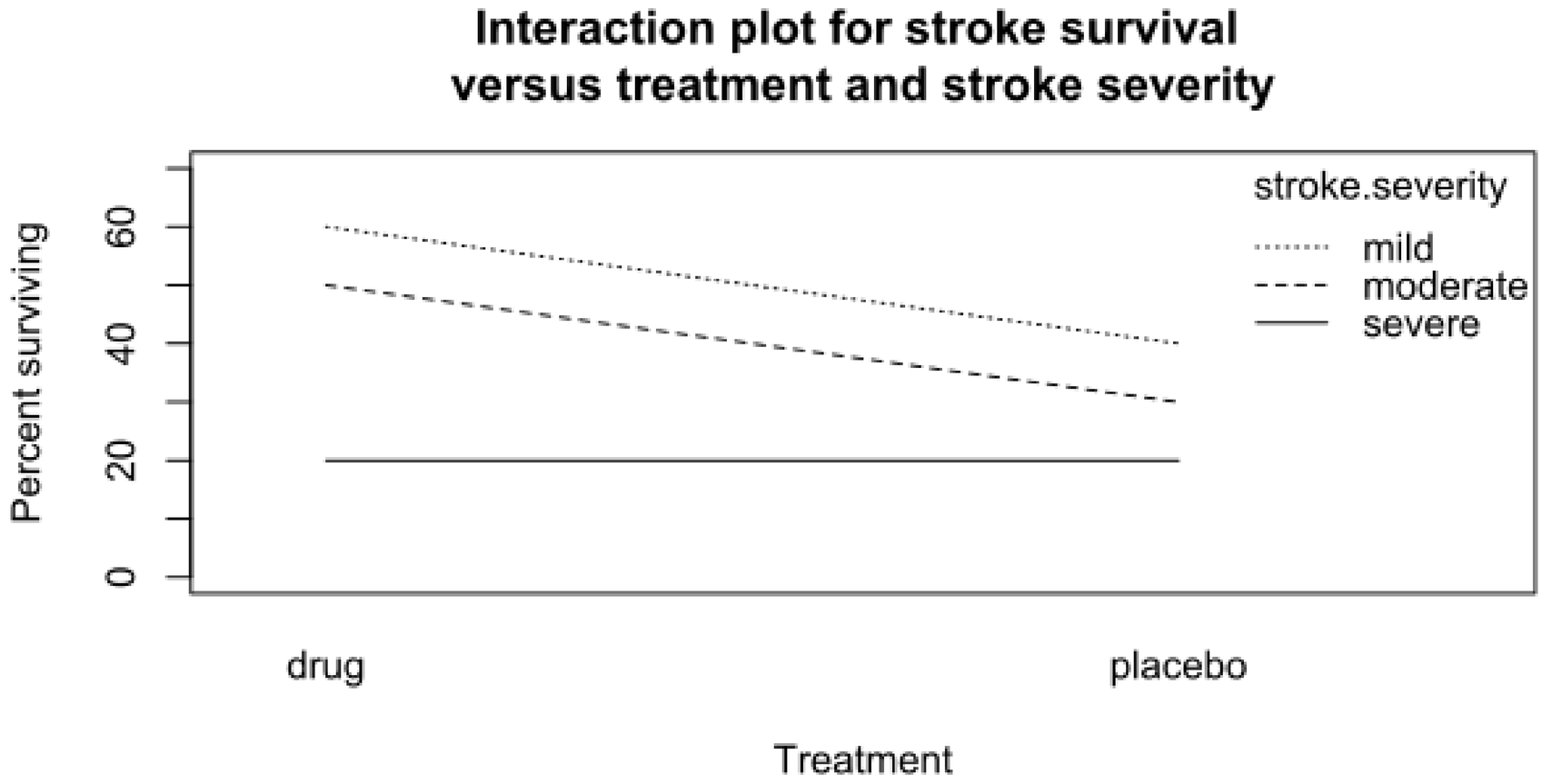
- Effect of an IV varies across groups
- The relationship between IV and DV is non-linear (i.e. quadratic)
- Variables may have non-normal distribution
- Two or more IVs are correlated (multicollinearity)

INTERACTION TERMS

Effect of one IV depends on the other IV:



INTERACTION TERMS



INTERACTION TERMS

Interpretation

- **The effect is not on DV but on the link between DV and another IV**
- **Also depends on the scales of the IVs (categorical or continuous)**
- **3 types of IT:**

categorical*categorical,
categorical*continuous,
continuous*continuous

HOW TO ESTIMATE INTERACTION TERMS IT IN R

Data from WVS wave 7 in Russia (ruswvs7 in LM3.RData)
Variables:

- age1 = age in years
- income = income level (self-placement, 1 – lowest level, 10 – highest level)
- sex = gender (female, male)
- edu = level of education (categorical, primary, sec&prof, tertiary
- eduf = level of education (numeric, 1-9)
- H_URBRURAL = area of residence (unban, rural)
- sat = life satisfaction (1 – completely dissatisfied, 10 – completely satisfied)

HOW TO ESTIMATE INTERACTION TERMS IT IN R

```
> model0 = lm(sat~age1+sex+eduf+sex+H_URBRURAL+eduf+income, data = ruswvs7)
> summary(model0)

Call:
lm(formula = sat ~ age1 + sex + eduf + sex + H_URBRURAL + eduf +
    income, data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9823 -1.3082  0.0367  1.2911  4.8177

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.562435   0.269179  20.664  < 2e-16 ***
age1           -0.009518   0.002937  -3.241  0.00121 **
sexMale         0.059918   0.098067   0.611  0.54129
eduf           -0.030218   0.028300  -1.068  0.28577
H_URBRURALRural -0.054334   0.109268  -0.497  0.61907
income          0.327980   0.026454  12.398  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.965 on 1720 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.1099,    Adjusted R-squared:  0.1073
F-statistic: 42.48 on 5 and 1720 DF,  p-value: < 2.2e-16
```

HOW TO ESTIMATE INTERACTION TERMS IT IN R

```
> model1 = lm(sat~age1+sex*eduf+H_URBRURAL+income, data = ruswvs7)
> summary(model1)

Call:
lm(formula = sat ~ age1 + sex * eduf + sex + H_URBRURAL + eduf +
    income, data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1583 -1.3183  0.0406  1.3042  4.9896

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.889871   0.299913  19.639  < 2e-16 ***
age1            -0.010049   0.002941  -3.417 0.000647 ***
sexMale         -0.718625   0.331166  -2.170 0.030145 *
eduf            -0.081644   0.035145  -2.323 0.020294 *
H_URBRURALRural -0.042483   0.109214  -0.389 0.697330
income           0.327590   0.026416  12.401  < 2e-16 ***
sexMale:eduf     0.133205   0.054127   2.461 0.013954 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.962 on 1719 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.113,     Adjusted R-squared:  0.1099
F-statistic: 36.51 on 6 and 1719 DF,  p-value: < 2.2e-16
```

HOW TO ESTIMATE INTERACTION TERMS IT IN R

```
> model1 = lm(sat~age1+sex*eduf+H_URBRURAL+income, data = ruswvs7)
> summary(model1)
```

Interaction term

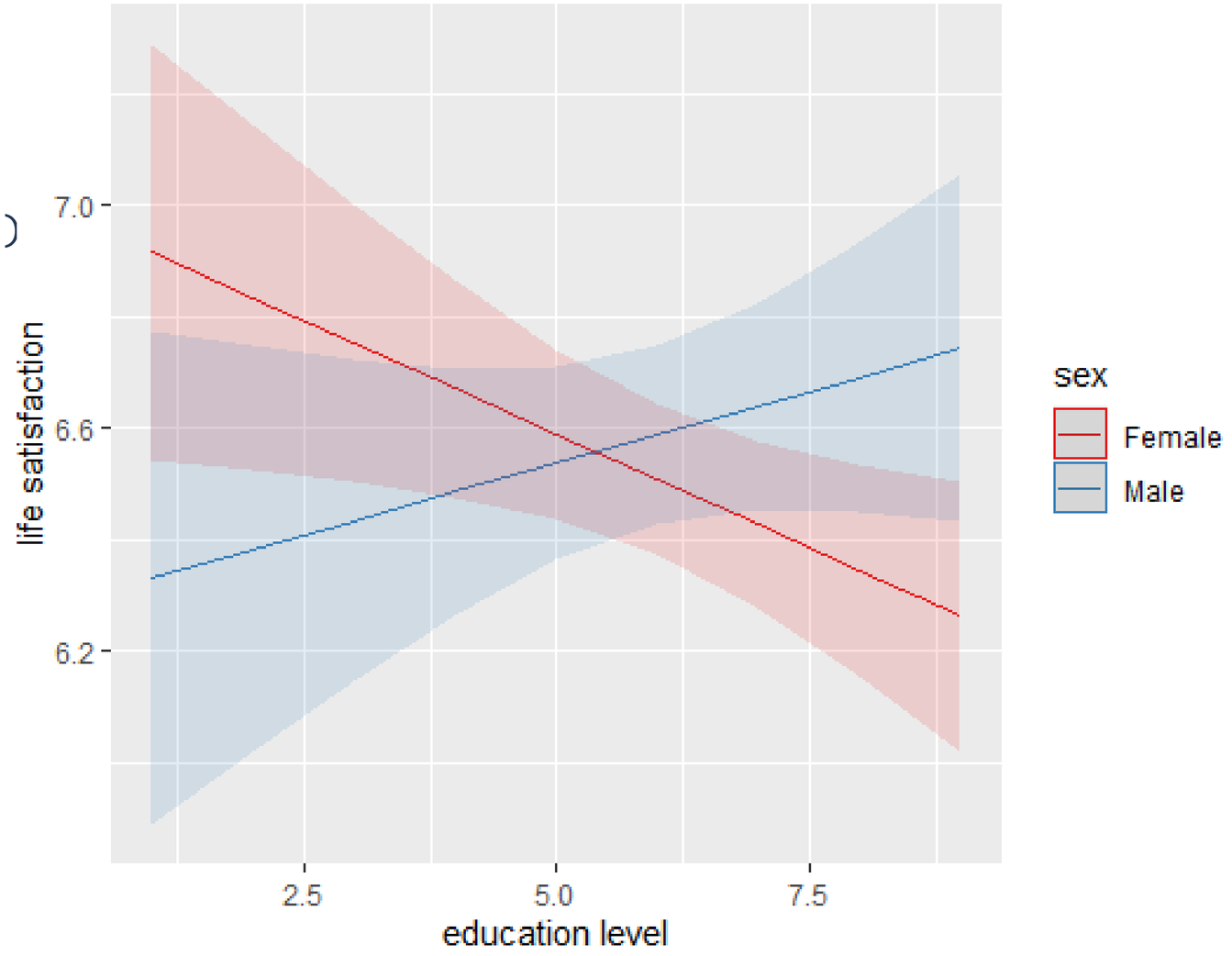
```
Call:
lm(formula = sat ~ age1 + sex * eduf + sex + H_URBRURAL + eduf +
    income, data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1583 -1.3183  0.0406  1.3042  4.9896

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.889871   0.299913  19.639  < 2e-16 ***
age1            -0.010049   0.002941  -3.417 0.000647 ***
sexMale         -0.718625   0.331166  -2.170 0.030145 *
eduf            -0.081644   0.035145  -2.323 0.020294 *
H_URBRURALRural -0.042483   0.109214  -0.389 0.697330
income           0.327590   0.026416  12.401  < 2e-16 ***
sexMale:eduf     0.133205   0.054127   2.461 0.013954 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.962 on 1719 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.113,     Adjusted R-squared:  0.1099
F-statistic: 36.51 on 6 and 1719 DF,  p-value: < 2.2e-16
```

HOW TO ESTIMATE INTERACTION TERMS IT IN R

```
> model1 = lm(sat~age1+sex*eduf+H_URBRURAL+income, data = ruswvs7)
> summary(model1)

Call:
lm(formula = sat ~ age1 + sex * eduf + sex + H_URBRURAL + eduf +
    income, data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1583 -1.3183  0.0406  1.3042  4.9896

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.889871   0.299913  19.639  < 2e-1
age1            -0.010049   0.002941  -3.417 0.0006     **
sexMale         -0.718625   0.331166  -2.170 0.030      *
eduf            -0.081644   0.035145  -2.323 0.02       4 *
H_URBRURALRural -0.042483   0.109214  -0.389 0.6  7330
income           0.327590   0.026416  12.401  < 2e-16 ***
sexMale:eduf     0.133205   0.054127   2.461 0.013954 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.962 on 1719 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.113,     Adjusted R-squared:  0.1099
F-statistic: 36.51 on 6 and 1719 DF,  p-value: < 2.2e-16
```

Education works differently for men and women

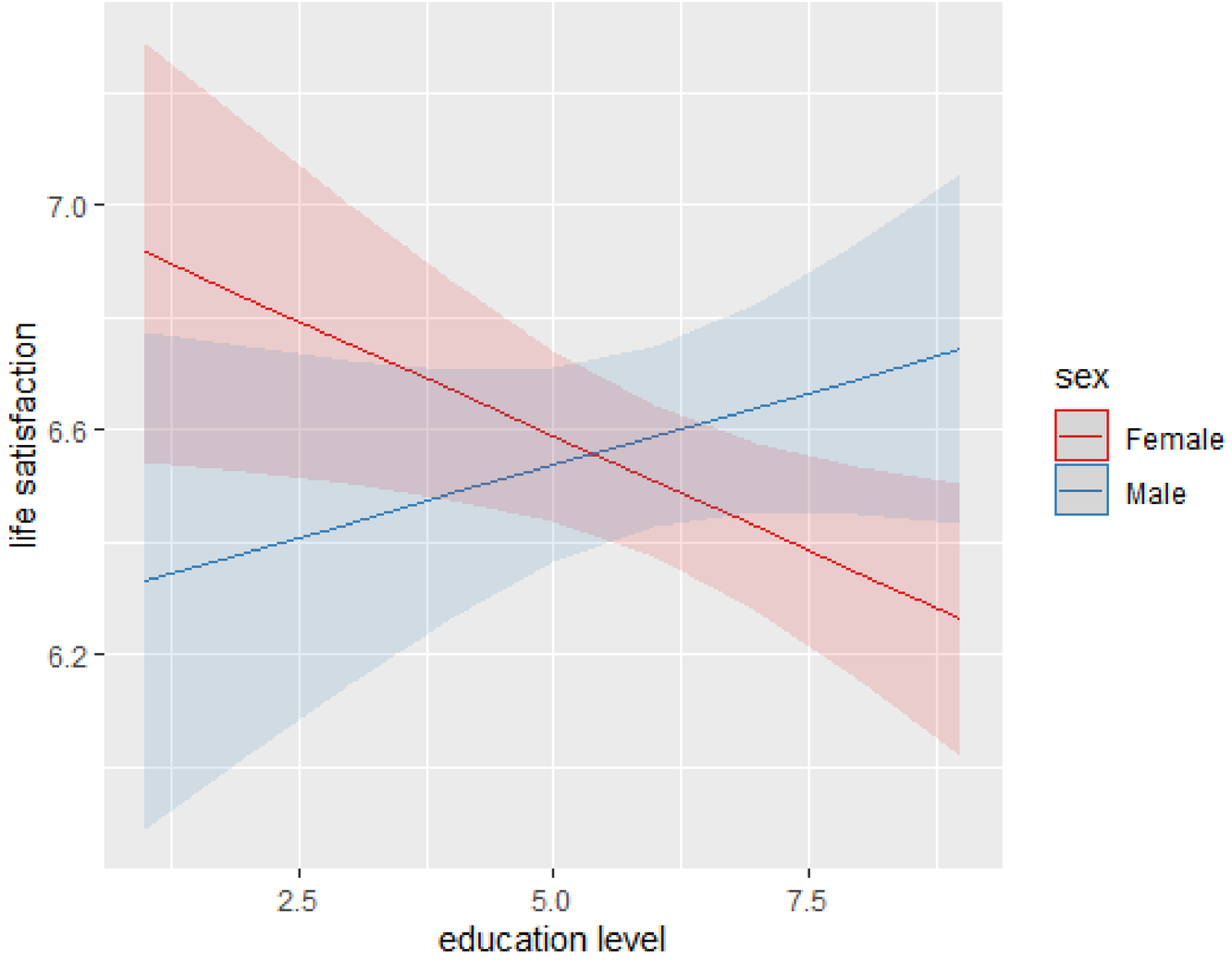
```
plot_model(model1, type = "pred",
           terms=c("eduf", "sex"),
           title = "",
           axis.title = c("education level",
                          "life satisfaction"))
```



```
plot_model(model1, type = "pred",
           terms=c("eduf", "sex"),
           title = "",
           axis.title = c("education level",
                          "life satisfaction"))
```

For women the link is negative, for men the link is positive

Women with higher level of education have lower life satisfaction, whereas men with higher level of education are more satisfied with their lives



HOW TO ESTIMATE INTERACTION TERMS IT IN R

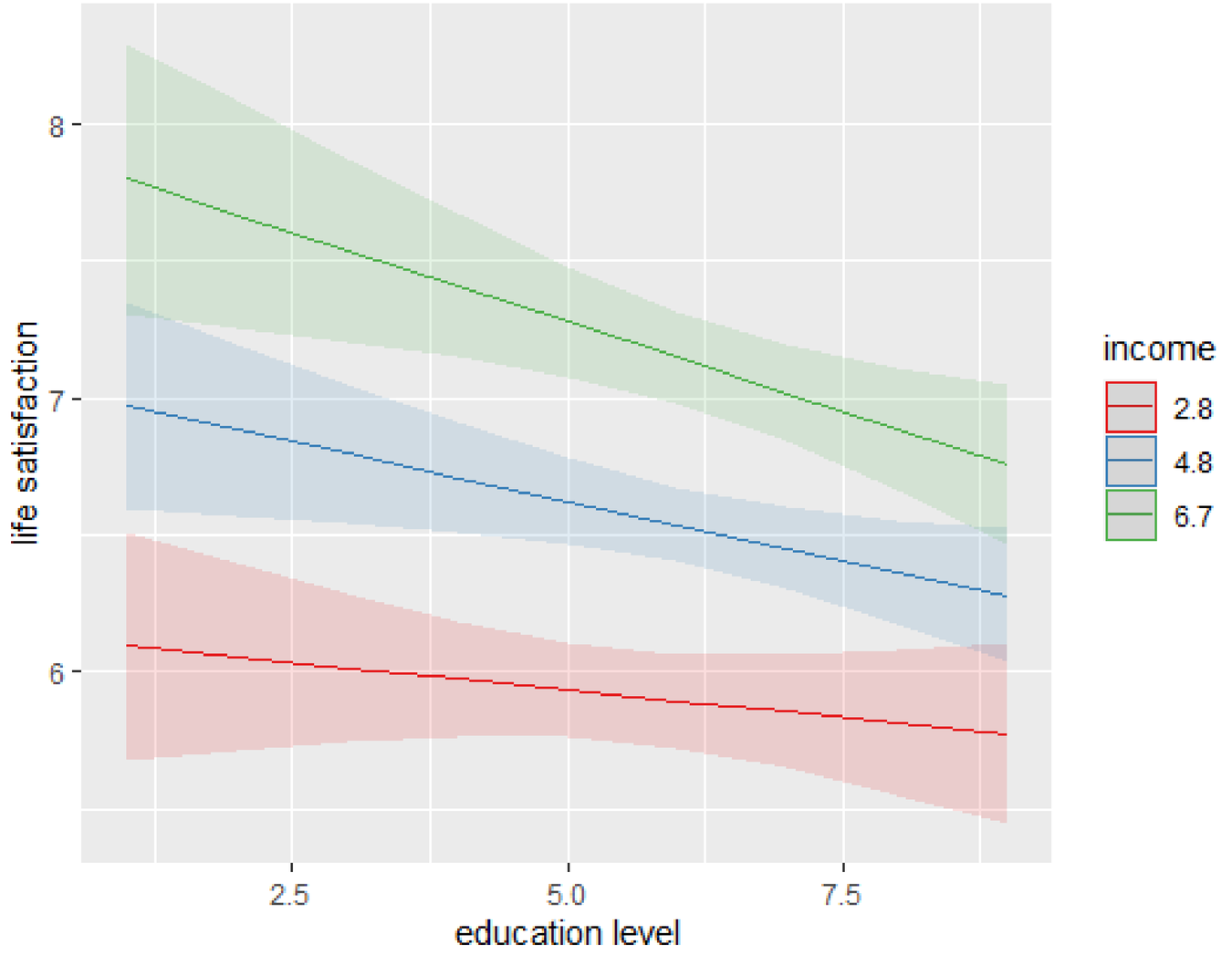
```
> model2 = lm(sat~age1+sex*eduf+H_URBRURAL+eduf*income, data = ruswvs7)
> summary(model2)

Call:
lm(formula = sat ~ age1 + sex * eduf + H_URBRURAL + eduf * income,
    data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6656 -1.3293  0.0332  1.2873  5.2106

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.280658   0.460452  11.468  < 2e-16 ***
age1           -0.009558   0.002952  -3.238  0.00123 **
sexMale        -0.772821   0.332427  -2.325  0.02020 *
eduf            0.025036   0.070570   0.355  0.72280
H_URBRURALRural -0.040269   0.109157  -0.369  0.71224
income          0.460069   0.080464   5.718 1.27e-08 ***
sexMale:eduf    0.142514   0.054358   2.622  0.00883 **
eduf:income    -0.023144   0.013279  -1.743  0.08153 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.961 on 1718 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.1146,    Adjusted R-squared:  0.111
F-statistic: 31.77 on 7 and 1718 DF,  p-value: < 2.2e-16
```

HOW TO ESTIMATE INTERACTION TERMS IT IN R

```
> model2 = lm(sat~age1+sex*eduf+H_URBRURAL+eduf*income, data = ruswvs7)
> summary(model2)

Call:
lm(formula = sat ~ age1 + sex * eduf + H_URBRURAL + eduf * income,
    data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6656 -1.3293  0.0332  1.2873  5.2106

Coefficients:
                 Estimate Std. Error t value P
(Intercept)      5.280658   0.460452  11.468
age1            -0.009558   0.002952  -3.238
sexMale         -0.772821   0.332427  -2.325
eduf             0.025036   0.070570   0.355  0.722
H_URBRURALRural -0.040269   0.109157  -0.369  0.71
income           0.460069   0.080464   5.718 1.27e    ***
sexMale:eduf     0.142514   0.054358   2.622  0.00883 **
eduf:income     -0.023144   0.013279  -1.743  0.08153 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.961 on 1718 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.1146,    Adjusted R-squared:  0.111
F-statistic: 31.77 on 7 and 1718 DF,  p-value: < 2.2e-16
```

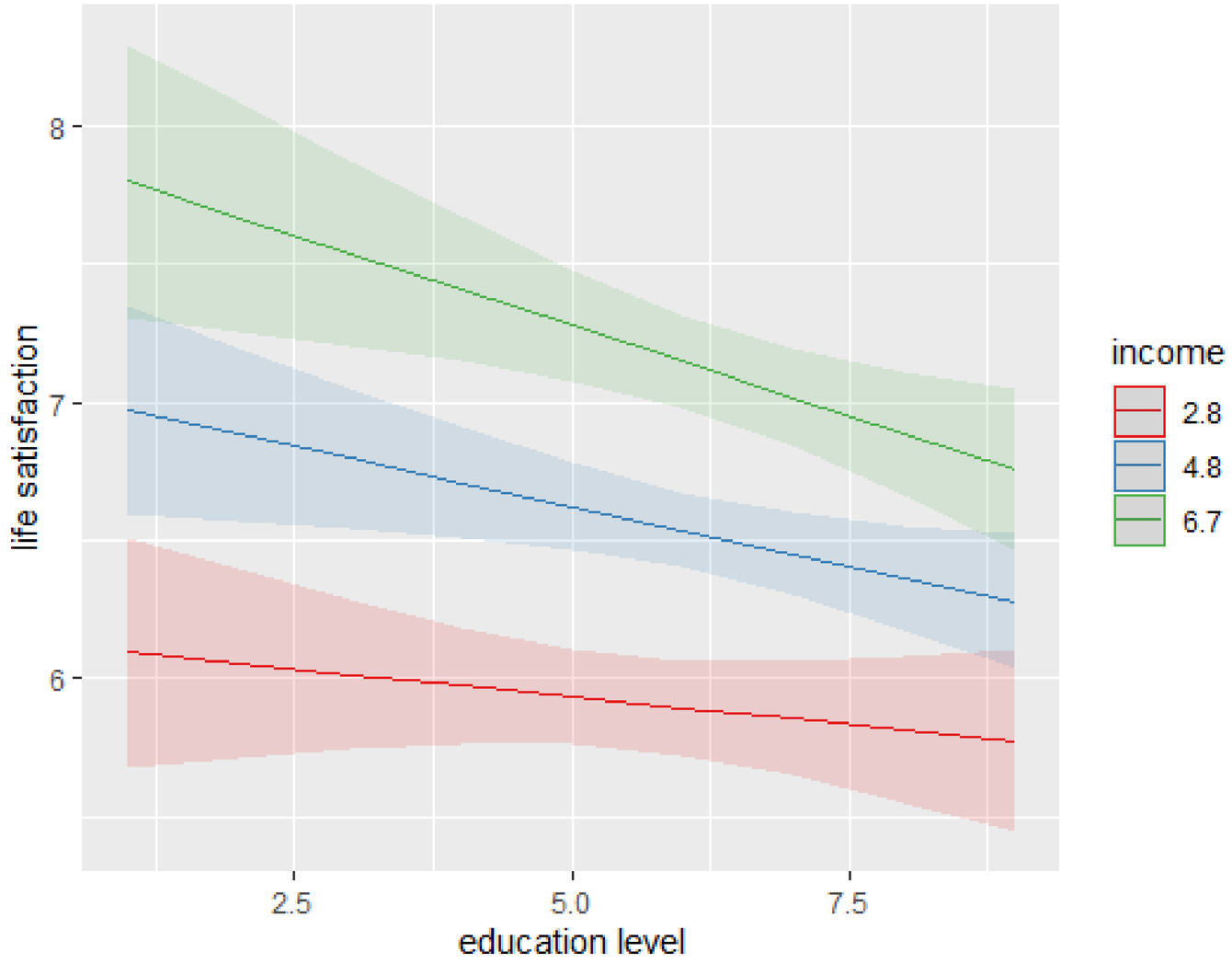
Marginally significant, let’s have a closer look

```
plot_model(model2, type = "pred",
           terms=c("eduf", “income"),
           title = "",
           axis.title = c("education level",
                          "life satisfaction"))
```



```
plot_model(model2, type = "pred",
           terms=c("eduf", “income"),
           title = "",
           axis.title = c("education level",
                          "life satisfaction"))
```

The impact of income on life satisfaction weakens as education level grows



HOW TO ESTIMATE INTERACTION TERMS IT IN R

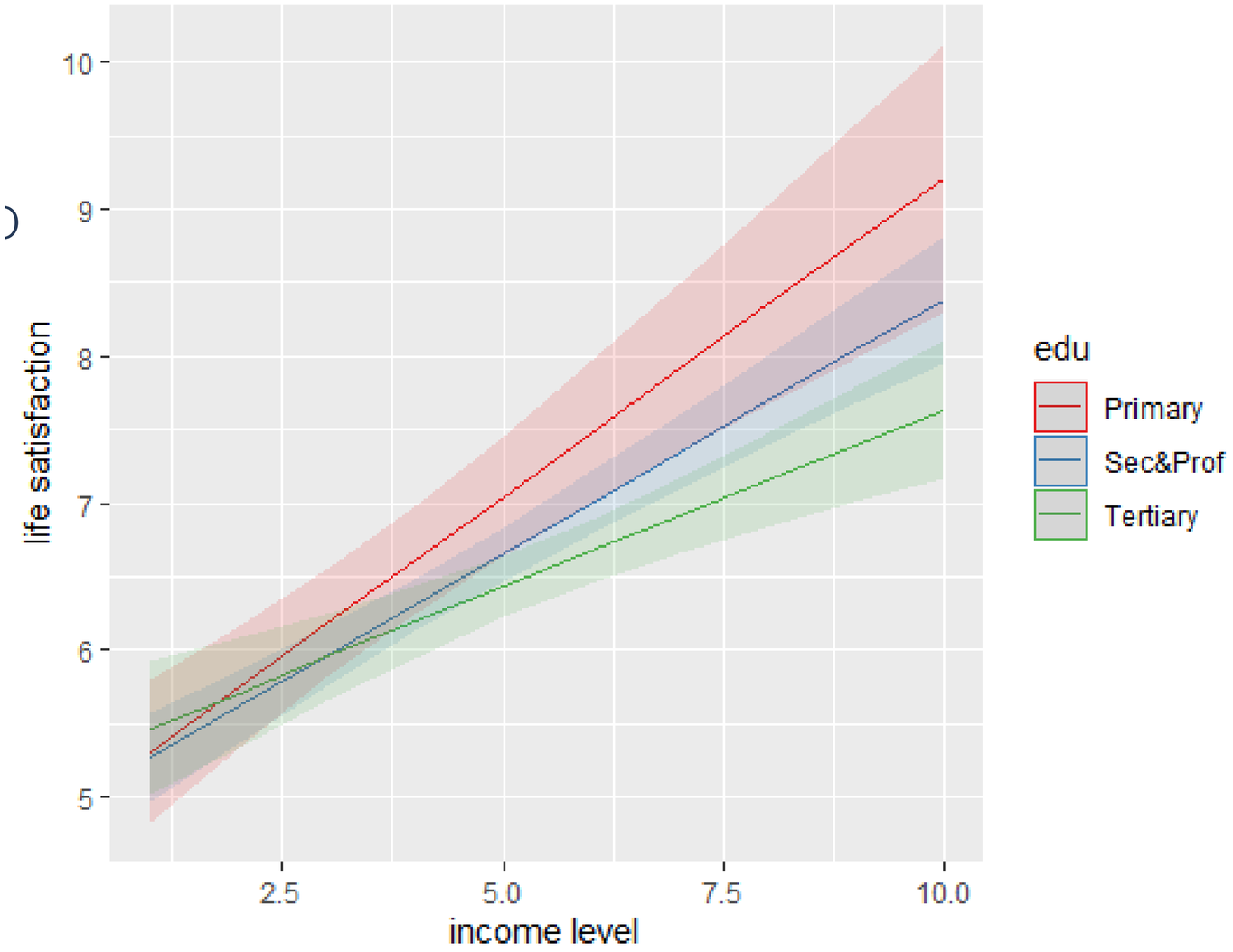
```
> model2.1 = lm(sat~age1+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model2.1)

Call:
lm(formula = sat ~ age1 + sex * edu + H_URBRURAL + edu * income,
    data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.8292 -1.3326  0.0337  1.2708  5.2986

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.331836   0.353593  15.079  < 2e-16 ***
age1               -0.009963   0.002985  -3.338 0.000862 ***
sexMale            -0.426193   0.270070  -1.578 0.114732
eduSec&Prof         0.045638   0.339105   0.135 0.892957
eduTertiary         0.348347   0.402156   0.866 0.386501
H_URBRURALRural    -0.039778   0.108626  -0.366 0.714268
income              0.433427   0.066073   6.560 7.11e-11 ***
sexMale:eduSec&Prof 0.366960   0.299204   1.226 0.220195
sexMale:eduTertiary 0.879246   0.319913   2.748 0.006052 **
eduSec&Prof:income -0.087184   0.074683  -1.167 0.243214
eduTertiary:income -0.192222   0.080978  -2.374 0.017718 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.96 on 1715 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.1171,    Adjusted R-squared:  0.1119
```

Let’s look at the categories of education

HOW TO ESTIMATE INTERACTION TERMS IT IN R

```
> model2.1 = lm(sat~age1+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model2.1)

Call:
lm(formula = sat ~ age1 + sex * edu + H_URBRURAL + edu * income,
    data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-7.8292 -1.3326  0.0337  1.2708  5.2986

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.331836   0.353593  15.079  < 2e-16 ***
age1               -0.009963   0.002985  -3.338 0.000862 ***
sexMale            -0.426193   0.270070  -1.578 0.114732
eduSec&Prof         0.045638   0.339105   0.135 0.892957
eduTertiary         0.348347   0.402156   0.866 0.386501
H_URBRURALRural    -0.039778   0.108626  -0.366 0.714268
income              0.433427   0.066073   6.560 7.11e-11 ***
sexMale:eduSec&Prof 0.366960   0.299204   1.226 0.220195
sexMale:eduTertiary 0.879246   0.319913   2.748 0.006052 **
eduSec&Prof:income -0.087184   0.074683  -1.167 0.243214
eduTertiary:income -0.192222   0.080978  -2.374 0.017718 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.96 on 1715 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.1171,    Adjusted R-squared:  0.1119
```

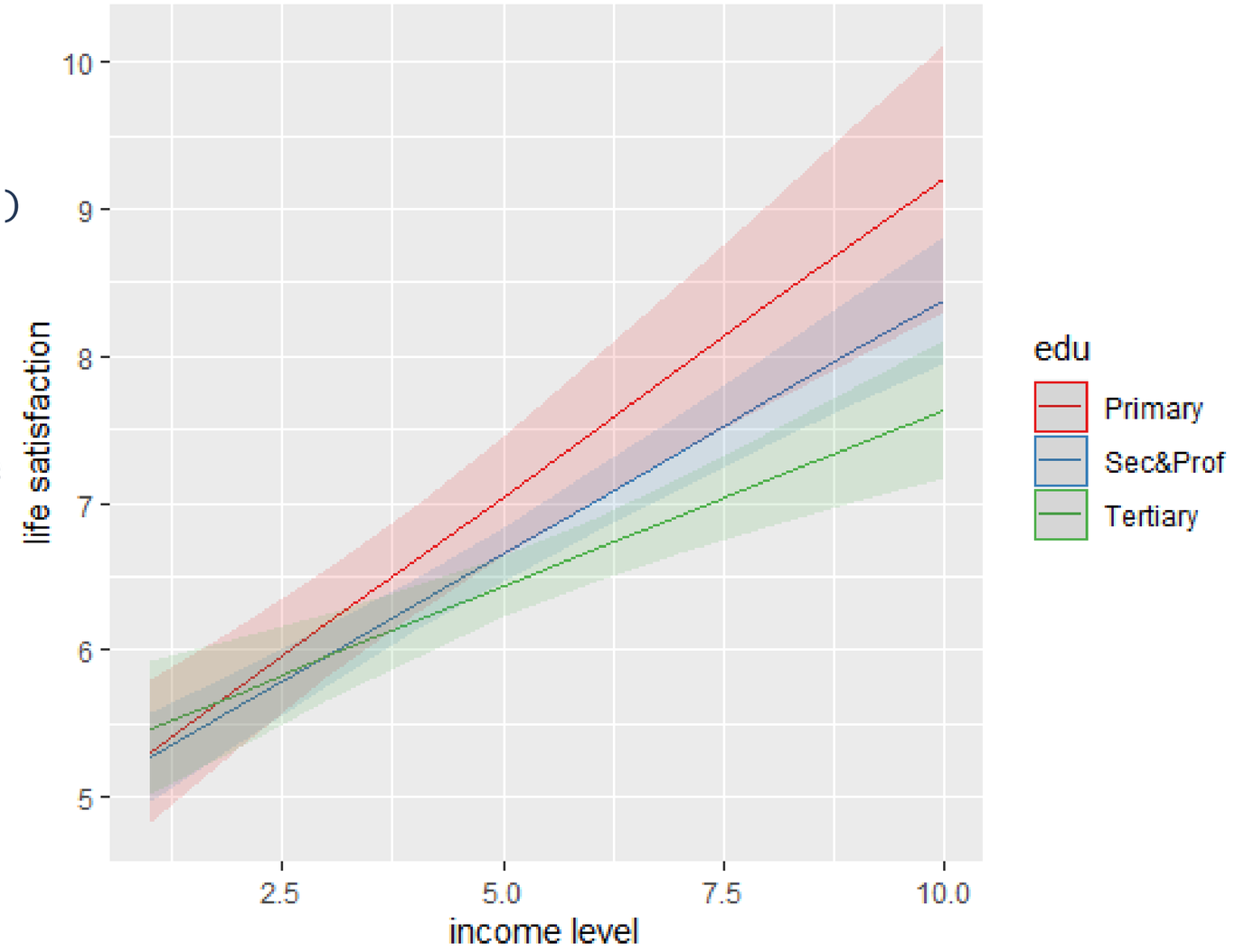
```
plot_model(model2.1, type = "pred",
           terms=c("income", "edu"),
           title = "",
           axis.title = c("incomelevel",
                          "life satisfaction"))
```



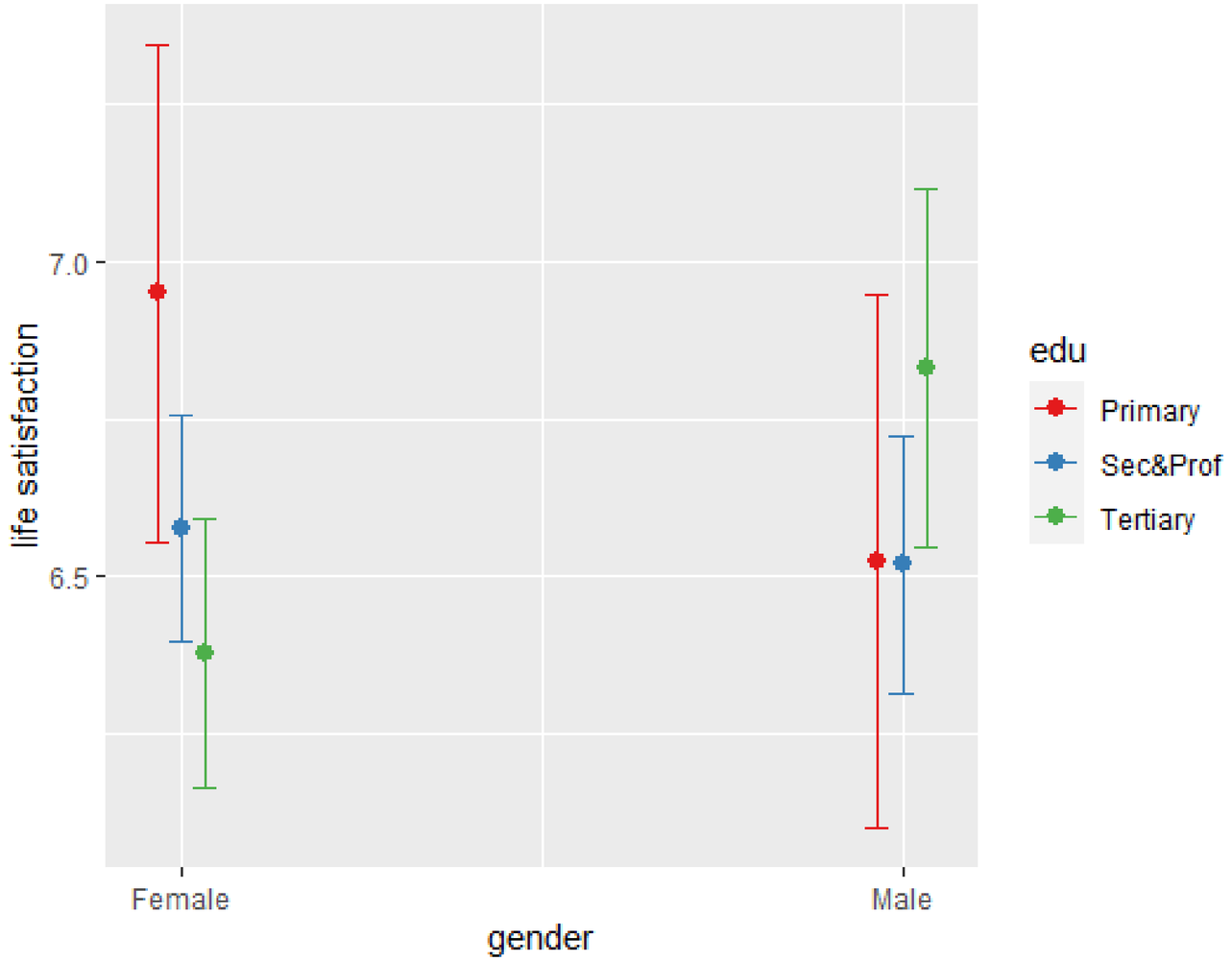
```
plot_model(model2.1, type = "pred",
           terms=c("income", "edu"),
           title = "",
           axis.title = c("incomelevel",
                          "life satisfaction"))
```

For people with primary education the effect of income on life satisfaction is higher

As income level grows, the difference in life satisfaction among education groups becomes more prominent

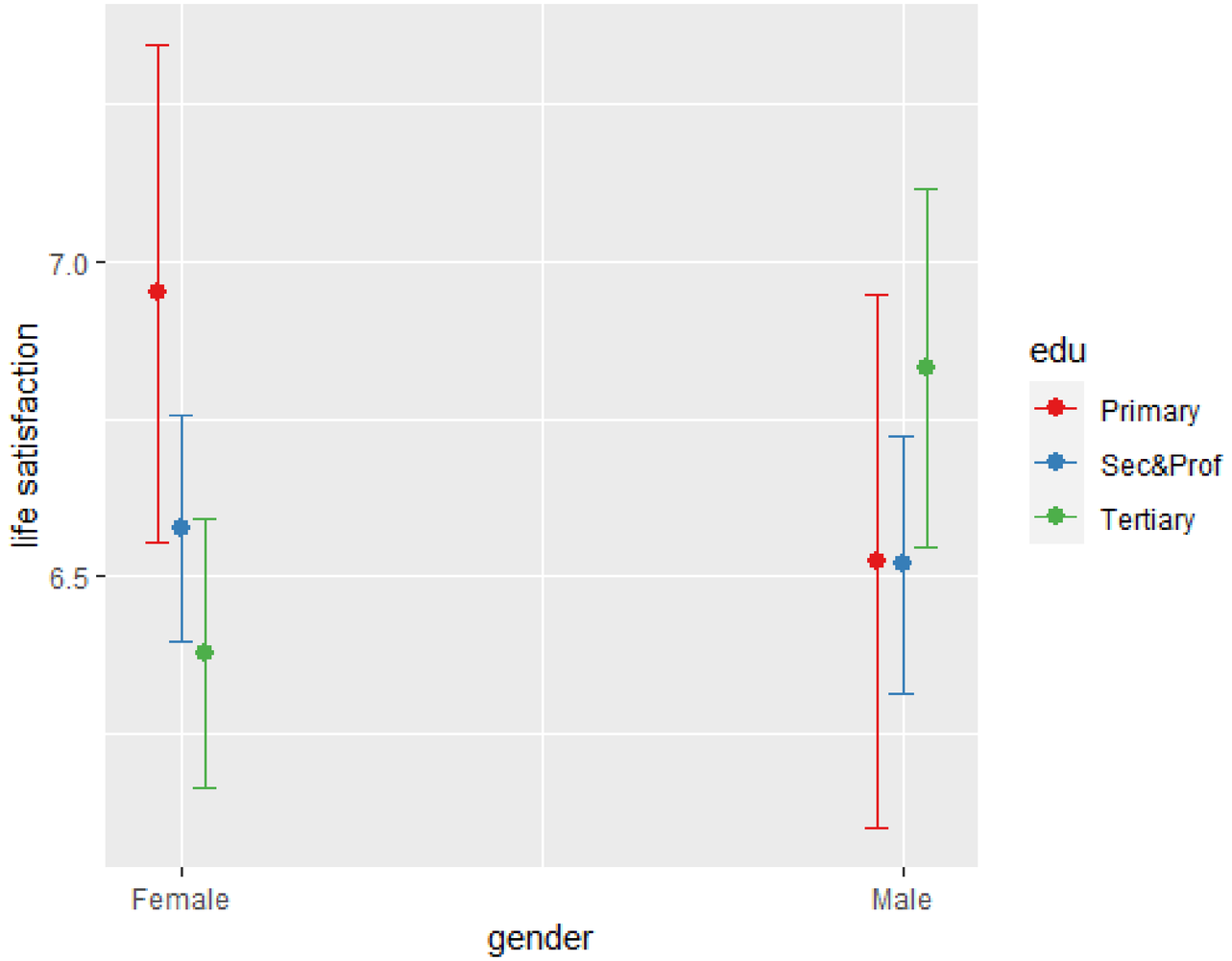


```
plot_model(model2.1, type = "pred",
           terms=c("sex", "edu"),
           title = "",
           axis.title = c("gender",
                          "life satisfaction"))
```



```
plot_model(model2.1, type = "pred",
           terms=c("sex", "edu"),
           title = "",
           axis.title = c("gender",
                          "life satisfaction"))
```

Men with tertiary education have higher life satisfaction than women with tertiary education



TRANSFORMATION OF VARIABLES

- Effects may be non-linear
- Variables may be distributed non-normally

In this case it is possible to transform the variables

COMMON TYPES OF TRANSFORMATIONS

- Raising to the n^{th} power
- Log transformation
- Scaling (if the scales are considerably different)

RAISING TO THE POWER OF 2 (QUADRATIC TERMS)

```
> model3 = lm(sat~age1+I(age1^2)+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model3)

Call:
lm(formula = sat ~ age1 + I(age1^2) + sex * edu + H_URBRURAL + 
    edu * income, data = ruswvs7)
Residuals:
    Min      1Q  Median      3Q     Max 
-8.1594 -1.3675  0.0174  1.2702  5.3721 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          5.9830119  0.4459386  13.417  < 2e-16 ***
age1                -0.0474668  0.0159665  -2.973  0.00299 ** 
I(age1^2)            0.0003926  0.0001642   2.391  0.01691 *  
sexMale             -0.3647574  0.2709210  -1.346  0.17836    
eduSec&Prof          0.1887776  0.3438911   0.549  0.58311    
eduTertiary          0.5133028  0.4074868   1.260  0.20796    
H_URBRURALRural     -0.0298525  0.1085565  -0.275  0.78335    
income               0.4393251  0.0660285   6.654 3.84e-11 ***
sexMale:eduSec&Prof  0.3136172  0.2996249   1.047  0.29539    
sexMale:eduTertiary  0.8102063  0.3207761   2.526  0.01163 *  
eduSec&Prof:income  -0.0984022  0.0747279  -1.317  0.18808    
eduTertiary:income  -0.1985871  0.0809110  -2.454  0.01421 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.958 on 1714 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:   0.12,     Adjusted R-squared:  0.1144
```

RAISING TO THE POWER OF 2 (QUADRATIC TERMS)

```
> model3 = lm(sat~age1+I(age1^2)+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model3)

Call:
lm(formula = sat ~ age1 + I(age1^2) + sex * edu + H_URBRURAL +
    edu * income, data = ruswvs7)
Residuals:
    Min      1Q  Median      3Q     Max
-8.1594 -1.3675  0.0174  1.2702  5.3721

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          5.9830119  0.4459386  13.417  < 2e-16 ***
age1                -0.0474668  0.0159665  -2.973  0.00299 **
I(age1^2)            0.0003926  0.0001642   2.391  0.01691 *
sexMale             -0.3647574  0.2709210  -1.346  0.17836
eduSec&Prof          0.1887776  0.3438911   0.549  0.58311
eduTertiary          0.5133028  0.4074868   1.260  0.20796
H_URBRURALRural     -0.0298525  0.1085565  -0.275  0.78335
income               0.4393251  0.0660285   6.654 3.84e-11 ***
sexMale:eduSec&Prof  0.3136172  0.2996249   1.047  0.29539
sexMale:eduTertiary  0.8102063  0.3207761   2.526  0.01163 *
eduSec&Prof:income  -0.0984022  0.0747279  -1.317  0.18808
eduTertiary:income  -0.1985871  0.0809110  -2.454  0.01421 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.958 on 1714 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:   0.12,     Adjusted R-squared:  0.1144
```

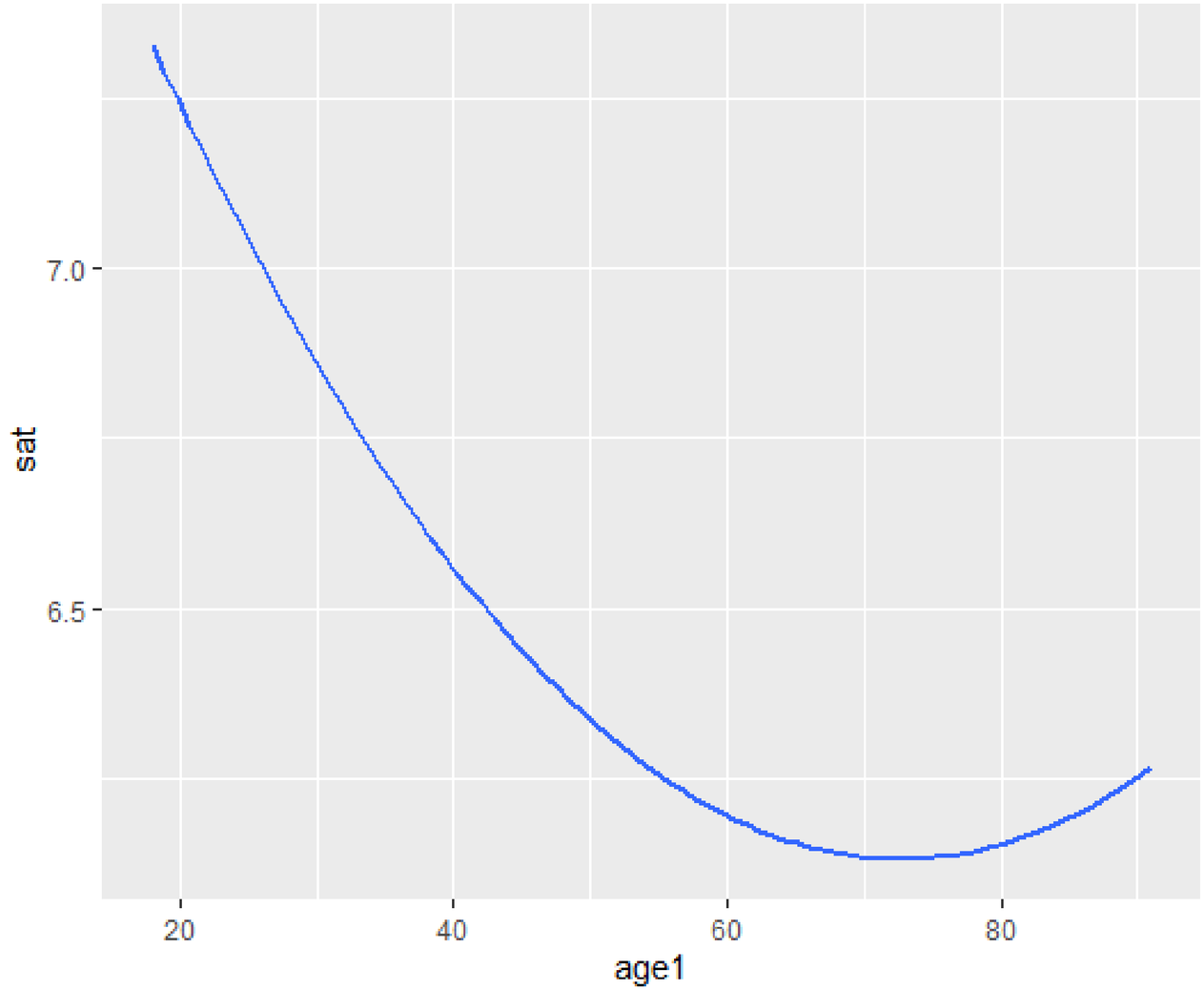
Quadratic term

RAISING TO THE POWER OF 2 (QUADRATIC TERMS)

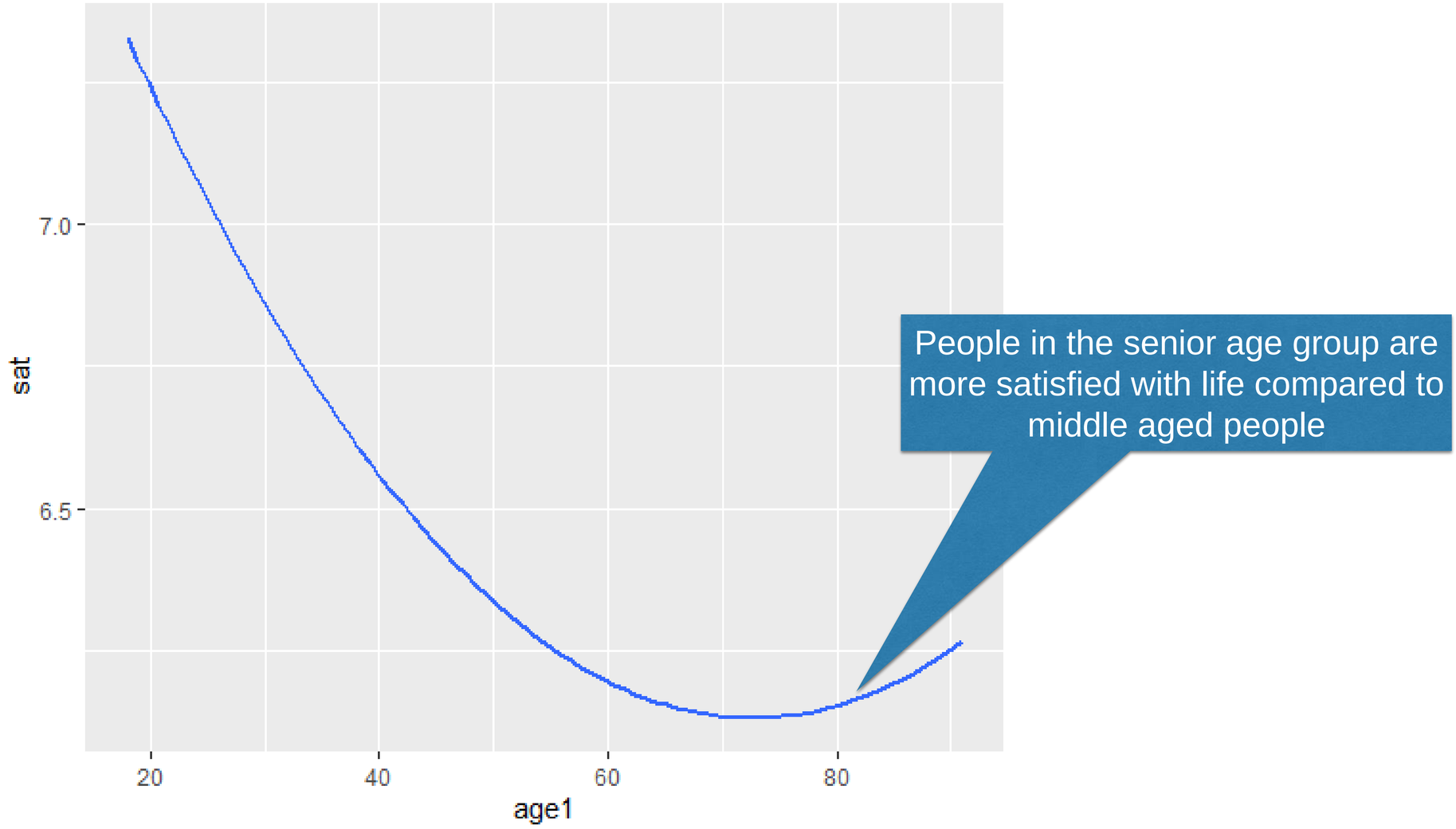
```
> model3 = lm(sat~age1+I(age1^2)+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model3)

Call:
lm(formula = sat ~ age1 + I(age1^2) + sex * edu + H_URBRURAL +
    edu * income, data = ruswvs7)
Residuals:
    Min      1Q  Median      3Q     Max
-8.1594 -1.3675  0.0174  1.2702  5.3721

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          5.9830119  0.4459386  13.417  < 2e-16 ***
age1                -0.0474668  0.0159665  -2.973  0.00299 **
I(age1^2)            0.0003926  0.0001642   2.391  0.01691 *
sexMale             -0.3647574  0.2709210  -1.346  0.17836
eduSec&Prof          0.1887776  0.3438911   0.549  0.58311
eduTertiary          0.5133028  0.4074868   1.260  0.20796
H_URBRURALRural     -0.0298525  0.1085565  -0.275  0.78335
income               0.4393251  0.0660285   6.654 3.84e-11 ***
sexMale:eduSec&Prof  0.3136172  0.2996249   1.047  0.29539
sexMale:eduTertiary  0.8102063  0.3207761   2.526  0.01163 *
eduSec&Prof:income  -0.0984022  0.0747279  -1.317  0.18808
eduTertiary:income  -0.1985871  0.0809110  -2.454  0.01421 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.958 on 1714 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:   0.12,     Adjusted R-squared:  0.1144
```

The link is indeed non-linear

```
ggplot(ruswvs7, aes(x=age1, y=sat)) + stat_smooth(se=F,method='lm', formula=y~x+I(x^2))
```



```
ggplot(ruswvs7, aes(x=age1, y=sat)) + stat_smooth(se=F,method='lm', formula=y~x+I(x^2))
```



SCALING

```
> model4 = lm(sat~scale(age1)+I(scale(age1)^2)+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model4)

Call:
lm(formula = sat ~ scale(age1) + I(scale(age1)^2) + sex * edu +
    H_URBRURAL + edu * income, data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-8.1594 -1.3675  0.0174  1.2702  5.3721

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          4.63720    0.31538  14.703  < 2e-16 ***
scale(age1)         -0.20225    0.05273  -3.836  0.00013 ***
I(scale(age1)^2)     0.11512    0.04815   2.391  0.01691 *
sexMale             -0.36476    0.27092  -1.346  0.17836
eduSec&Prof          0.18878    0.34389   0.549  0.58311
eduTertiary          0.51330    0.40749   1.260  0.20796
H_URBRURALRural     -0.02985    0.10856  -0.275  0.78335
income               0.43933    0.06603   6.654 3.84e-11 ***
sexMale:eduSec&Prof  0.31362    0.29962   1.047  0.29539
sexMale:eduTertiary  0.81021    0.32078   2.526  0.01163 *
eduSec&Prof:income  -0.09840    0.07473  -1.317  0.18808
eduTertiary:income  -0.19859    0.08091  -2.454  0.01421 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.958 on 1714 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:   0.12,     Adjusted R-squared:  0.1144
```

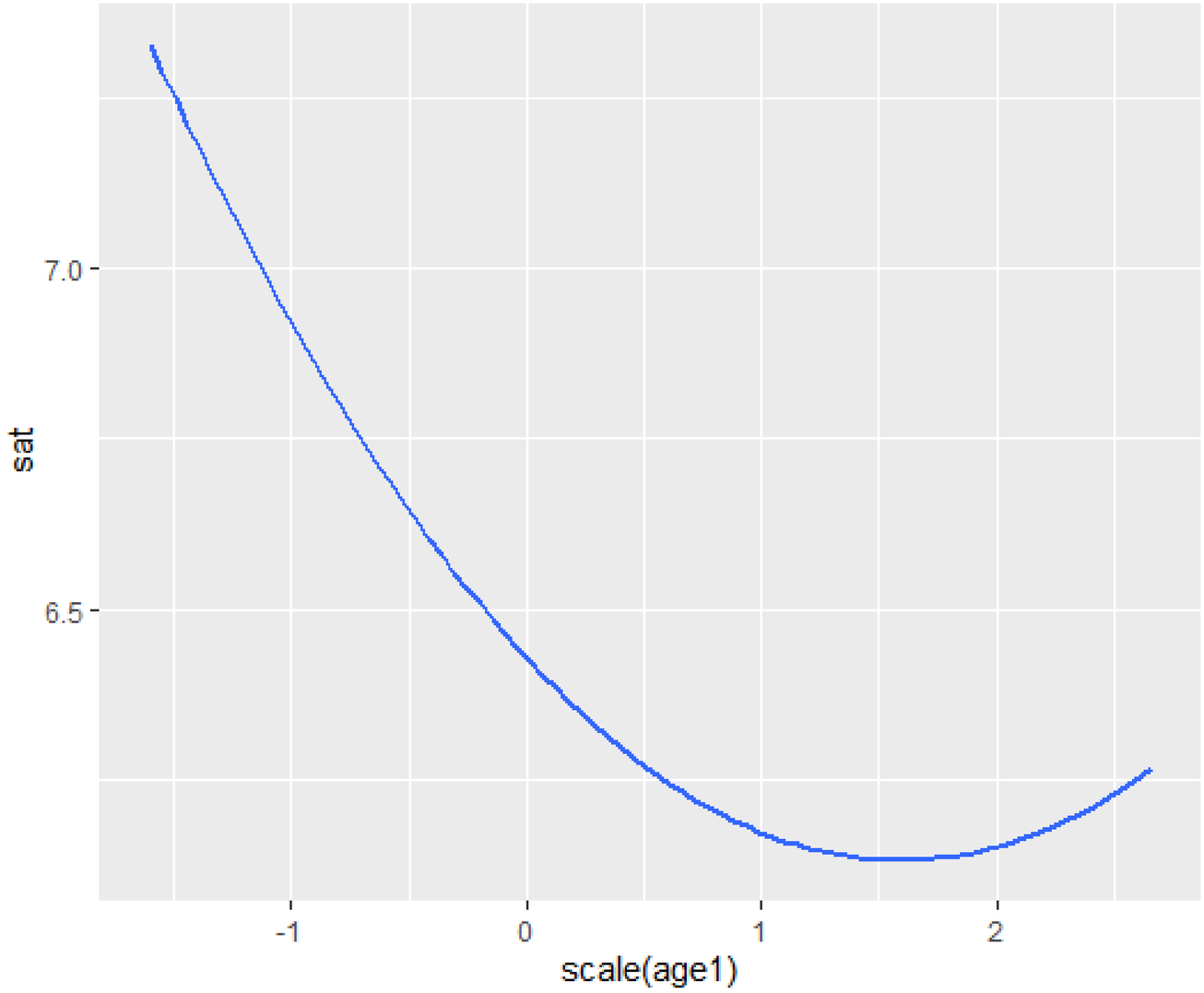
Scaling (standard deviations instead of values)

SCALING

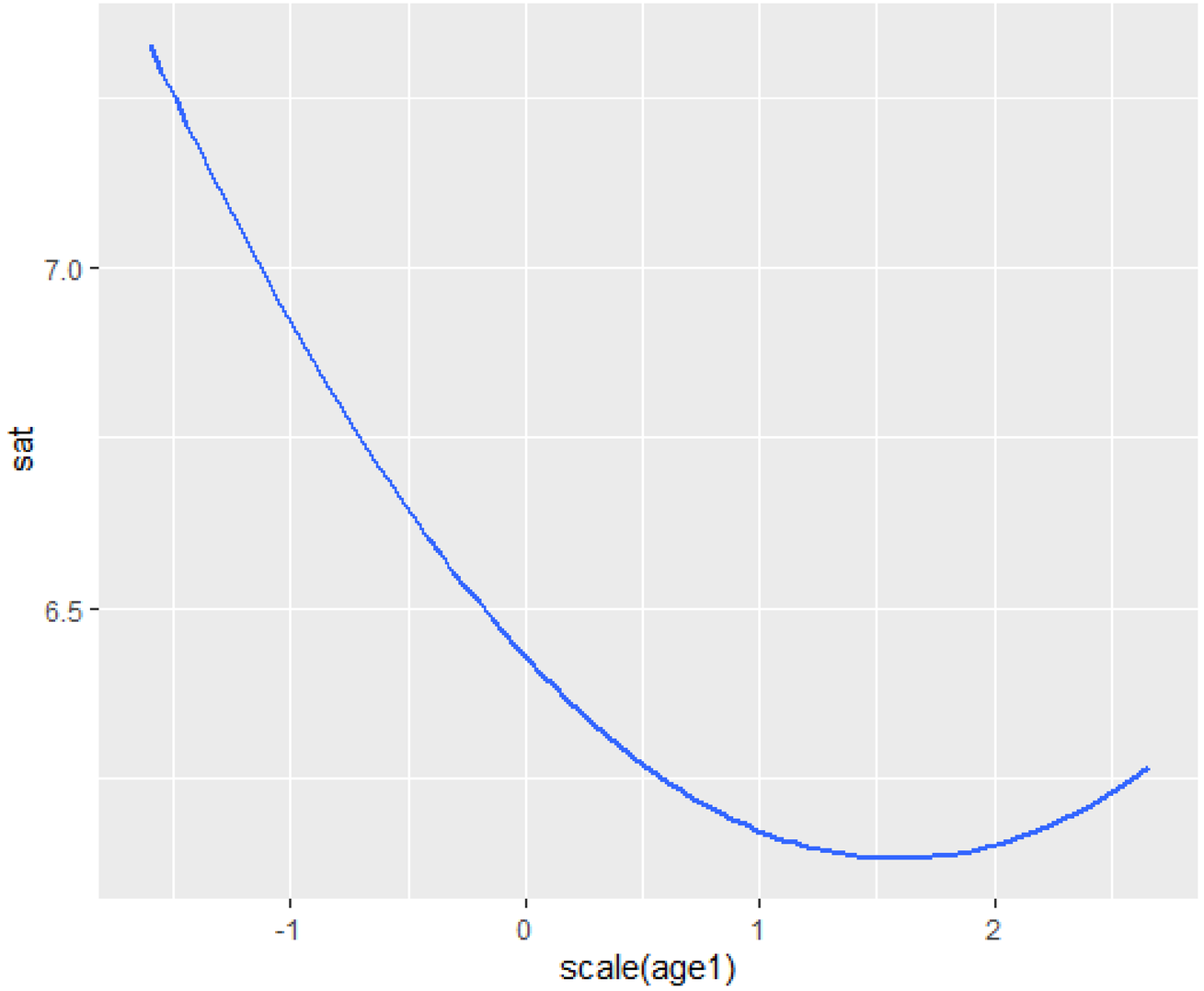
```
> model4 = lm(sat~scale(age1)+I(scale(age1)^2)+sex*edu+H_URBRURAL+edu*income, data = ruswvs7)
> summary(model4)

Call:
lm(formula = sat ~ scal                                    u +
    H_URBRURAL + edu *

Residuals:
    Min      1Q  Median              Max
-8.1594 -1.3675  0.0174  1.27        3721

Coefficients:
                   Estimate  d. Error t value Pr(>|t|)
(Intercept)         4.63720    0.31538  14.703  < 2e-16 ***
scale(age1)        -0.20225    0.05273  -3.836  0.00013 ***
I(scale(age1)^2)    0.11512    0.04815   2.391  0.01691 *
sexMale            -0.36476    0.27092  -1.346  0.17836
eduSec&Prof         0.18878    0.34389   0.549  0.58311
eduTertiary         0.51330    0.40749   1.260  0.20796
H_URBRURALRural    -0.02985    0.10856  -0.275  0.78335
income              0.43933    0.06603   6.654 3.84e-11 ***
sexMale:eduSec&Prof 0.31362    0.29962   1.047  0.29539
sexMale:eduTertiary 0.81021    0.32078   2.526  0.01163 *
eduSec&Prof:income -0.09840    0.07473  -1.317  0.18808
eduTertiary:income -0.19859    0.08091  -2.454  0.01421 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.958 on 1714 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:   0.12,     Adjusted R-squared:  0.1144
```

Scale is in st. dev. Increase in 1 st. dev. of age decreases life satisfaction by 0.2

```
ggplot(ruswvs7, aes(x=scale(age1), y=sat)) + stat_smooth(se=F,method='lm', formula=y~x+I(x^2))
```



```
ggplot(ruswvs7, aes(x=scale(age1), y=sat)) + stat_smooth(se=F,method='lm', formula=y~x+I(x^2))
```



Scale of age is not in years but in st.dev. Mean age is now 0, 1 refers to 1 standard deviation from the mean value

LOG TRANSFORMATION

```
> model5 = lm(sat~age1+sex+edu+H_URBRURAL+ log(income), data = ruswvs7)
> summary(model5)

Call:
lm(formula = sat ~ age1 + sex + edu + H_URBRURAL + log(income),
    data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3725 -1.3768  0.0258  1.2837  5.4056

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.290339   0.268732  19.686  < 2e-16 ***
age1            -0.009899   0.002977  -3.325 0.000902 ***
sexMale          0.067232   0.098332   0.684 0.494241
eduSec&Prof     -0.167735   0.153036  -1.096 0.273209
eduTertiary     -0.210854   0.170092  -1.240 0.215276
H_URBRURALRural -0.089365   0.108672  -0.822 0.410998
log(income)      1.265376   0.102512  12.344  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.966 on 1719 degrees of freedom
  (84 observations deleted due to missingness)
Multiple R-squared:  0.1098,    Adjusted R-squared:  0.1066
F-statistic: 35.32 on 6 and 1719 DF,  p-value: < 2.2e-16
```

Use the log of income

LOG TRANSFORMATION

```
> model5 = lm(sat~age1+sex+edu+H_URBRURAL+ log(income), data = ruswvs7)
> summary(model5)

Call:
lm(formula = sat ~ age1 + sex + edu + H_URBRURAL + log(income),
    data = ruswvs7)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3725 -1.3768  0.0258  1.2837  5.4056

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.290339   0.268732  19.686  < 2e-16 ***
age1            -0.009899   0.002977  -3.325 0.000902 ***
sexMale          0.067232   0.098332   0.684 0.494241
eduSec&Prof     -0.167735   0.153036  -1.096 0.273209
eduTertiary     -0.210854   0.170092  -1.240 0.215276
H_URBRURALRural -0.089365   0.108672  -0.822 0.410998
log(income)      1.265376   0.102512  12.344  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard er                                  f freedom
  (84 observations d
Multiple R-squared:                                   ed:  0.1066
F-statistic: 35.32 d                                  2.2e-16
```

The values are now in log

NATIONAL RESEARCH
UNIVERSITY