Interaction Models
(Part II)

Is there an effect of education on occupational prestige, controlling for occupation type and income level?

We will use the Prestige2.csv data. In these data, occupation type has three categories (bc, wc, and prof) and there are four fewer observations.

Occupational Type

Occupation type	N	p
Blue-collar	44	0.45
White-collar	23	0.32
Professional	31	0.23

Create dummy predictors

```
> Prestige$prof = ifelse(Prestige$type == "prof", 1, 0)
> Prestige$bc = ifelse(Prestige$type == "bc", 1, 0)
> Prestige$wc = ifelse(Prestige$type == "wc", 1, 0)
```

```
# Summary output for the dummy predictors

       bc               wc               prof       
 Min.   :0.000   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :0.000   Median :0.0000   Median :0.0000  
 Mean   :0.449   Mean   :0.2347   Mean   :0.3163  
 3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:1.0000  
 Max.   :1.000   Max.   :1.0000   Max.   :1.0000  
```

```
            prestige   education   L2income         bc          wc       prof
prestige   1.0000000  0.86647977  0.7512534 -0.6262996 -0.16554096  0.8207274
education  0.8664798  1.00000000  0.5904599 -0.8039530  0.04589092  0.8180538
L2income   0.7512534  0.59045990  1.0000000 -0.3320808 -0.26366233  0.5954640
bc        -0.6262996 -0.80395301 -0.3320808  1.0000000 -0.49987653 -0.6140064
wc        -0.1655410  0.04589092 -0.2636623 -0.4998765  1.00000000 -0.3766836
prof       0.8207274  0.81805379  0.5954640 -0.6140064 -0.37668362  1.0000000
```

Having a blue-collar occupation or a a white-collar occupation is negatively associated with prestige (i.e., lower average prestige for blue-collar and white-collar occupations)

Having a professional occupation is positively associated with education (i.e., higher average presitige for white-collar occupations)

Model

```
lm.bc = lm(prestige ~ 1 + L2income + wc + prof + education)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -81.2019    13.7431  -5.909 5.63e-08 ***
L2income      7.2694     1.1900   6.109 2.31e-08 ***
wc           -1.4394     2.3780  -0.605   0.5465
prof          6.7509     3.6185   1.866   0.0652 .
education     3.2845     0.6081   5.401 5.06e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.637 on 93 degrees of freedom
Multiple R-squared:  0.8555,	Adjusted R-squared:  0.8493
F-statistic: 137.6 on 4 and 93 DF,  p-value: < 2.2e-16
```

There is a statistically reliable effect of education on occupational prestige ($p < .001$), controlling for both income level, and occupation type.

Plotting the Fitted Values from the Model

For plotting, it is better to use the factor (type) than using dummy variables.

Model

```
lm.bc2 = lm(prestige ~ 1 + L2income + type + education)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -81.2019    13.7431  -5.909 5.63e-08 ***
L2income      7.2694     1.1900   6.109 2.31e-08 ***
typeprof      6.7509     3.6185   1.866   0.0652 .
typewc       -1.4394     2.3780  -0.605   0.5465
education     3.2845     0.6081   5.401 5.06e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.637 on 93 degrees of freedom
Multiple R-squared:  0.8555, Adjusted R-squared:  0.8493
F-statistic: 137.6 on 4 and 93 DF,  p-value: < 2.2e-16
```

When we use a factor, R dummy codes the predictors and chooses a reference group by default. From the output, we can identify the reference group because it is missing from the coefficients (bc).

We can use the `contrasts()` function to examine the contrast matrix without having to fit a model.

There are several different contrast matrices one could use, depending on what you want to test. The default in R is *treatment contrasts* (i.e., contrasts between the reference and other groups; dummy coding)

```
> contrasts(Prestige$type)

     prof wc
bc      0  0
prof    1  0
wc      0  1
```

The reference group has a row of all zeros (0 on each dummy); bc.

The first regression coefficient will represent the difference between the reference group (bc) and prof (1 on the first dummy).

The second regresssion coefficient will represent the difference between the reference group (bc) and wc (1 on the second dummy).

Set of a New Data Frame

Model

`prestige ~ 1 + L2income + type + education`

```
myData = expand.grid(
  education = seq(from = 6.3, to = 16, by = 0.1),
  L2income = mean(Prestige$L2income),
  type = c("bc", "wc", "prof")
  )
```

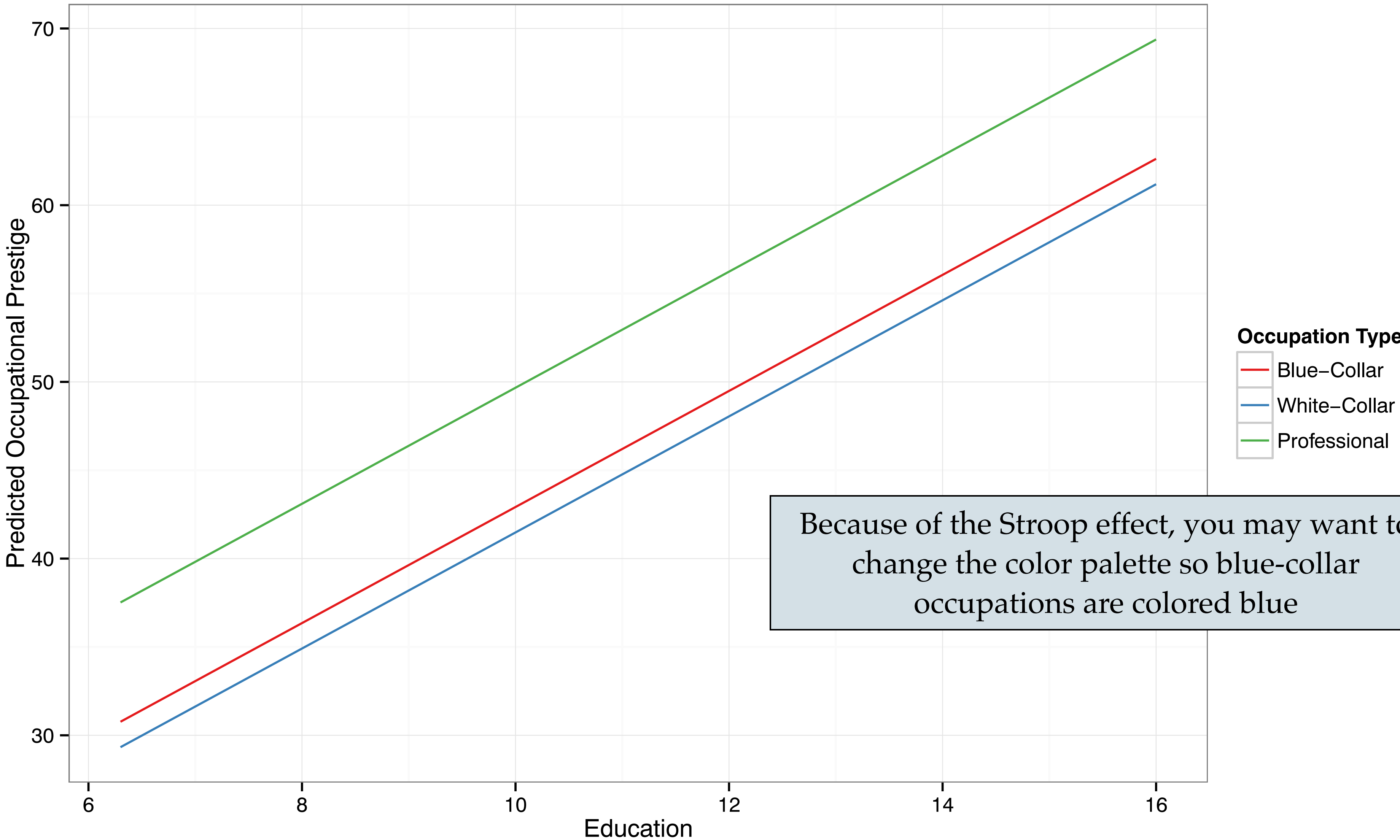
Predict and Bind the Data and Predictions

```
# Obtain fitted values
preds = predict(lm.bc2, newdata = myData)

# Bind the data and predicted values
myData = cbind(myData, preds)
```

Plot

```
ggplot(data = myData, aes(x = education, y = preds, color = type)) +
    geom_line() +
    xlab("Education") +
    ylab("Predicted Occupational Prestige") +
    scale_color_brewer(
      name = "Occupation Type",
      labels = c("Blue-Collar", "White-Collar", "Professional"),
      palette = "Set1") +
    theme_bw()
```

Note that we do not have to turn type into a factor...it already is a factor. Here we set the labels in the scale of the plot.



Because of the Stroop effect, you may want to change the color palette so blue-collar occupations are colored blue

Changing the Reference Group

We can also use the `contrasts()` function to assign a different reference group.

```
# Change the reference group and use dummy coding
> contrasts(Prestige$type) = contr.treatment(levels(Prestige$type),
     base = 2)
> contrasts(Prestige$type)

     bc wc
bc    1  0
prof  0  0
wc    0  1
```

contr.treatment() is the contrast matrix R uses for dummy coding

The argument base= sets the row to use as the reference group (here it will be `prof`)

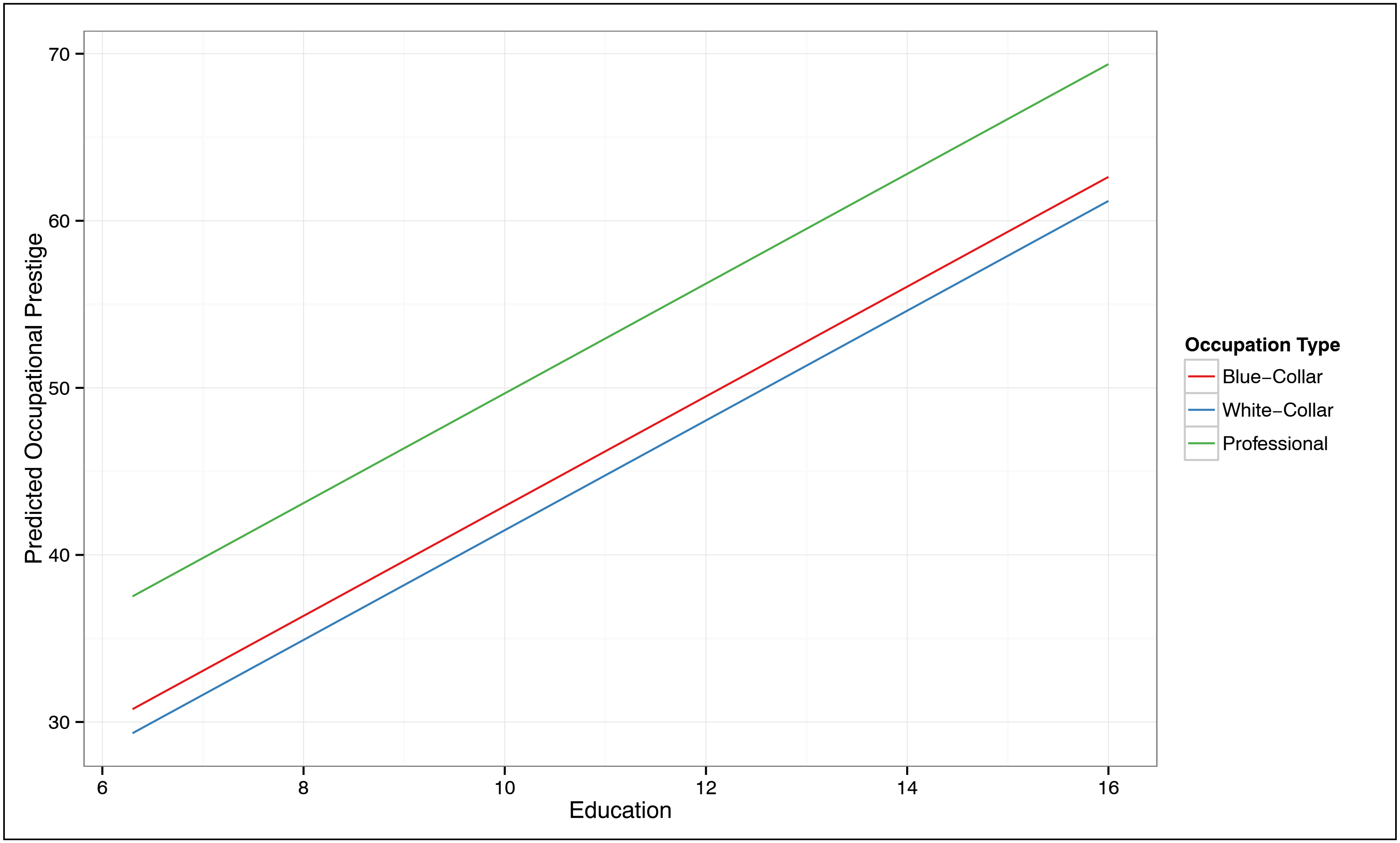
Model

`lm.bc = lm(prestige ~ 1 + L2income + type + education)`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -74.4510    15.1175  -4.925 3.65e-06 ***
L2income      7.2694     1.1900   6.109 2.31e-08 ***
typebc       -6.7509     3.6185  -1.866   0.0652 .
typewc       -8.1903     2.5882  -3.165   0.0021 **
education     3.2845     0.6081   5.401 5.06e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.637 on 93 degrees of freedom
Multiple R-squared:  0.8555, Adjusted R-squared:  0.8493
F-statistic: 137.6 on 4 and 93 DF,  p-value: < 2.2e-16
```

What do the coefficients mean when we use effects coding?

Regardless of reference group, the plot of the fitted regression lines are the same.



Changing the Contrasts Used

We can also use the `contrasts()` function assign different contrast matrices to be used in the regression. This is useful to test other types of contrast (e.g., Helmert, polynomial, etc.).

```
# Change the type of contrast matrix to effects coding
> contrasts(Prestige$type) = contr.sum(levels(Prestige$type))
> contrasts(Prestige$type)

     [,1] [,2]
bc      1    0
prof    0    1
wc     -1   -1
```

contr.sum() is the contrast matrix R uses for effects coding

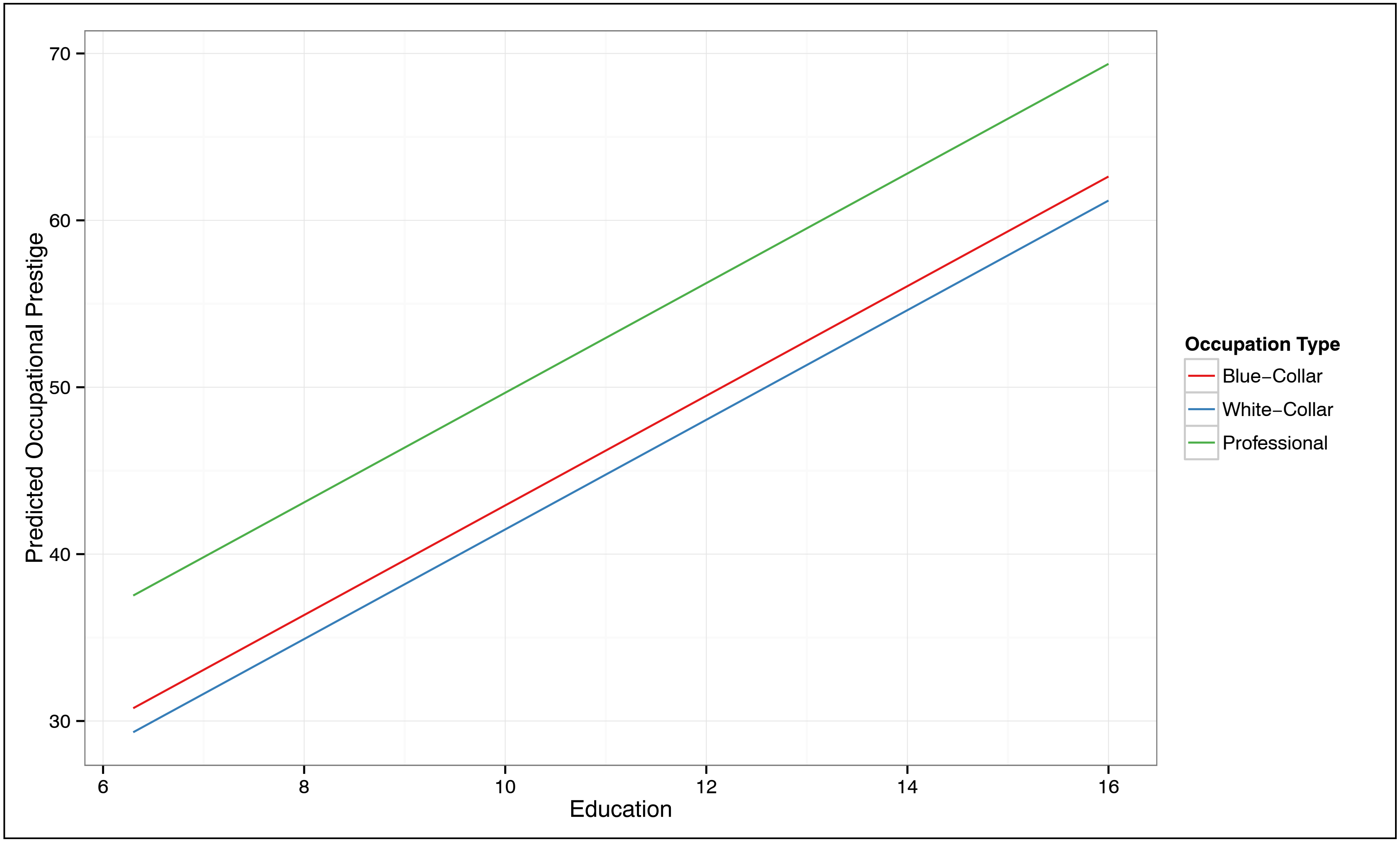
Model

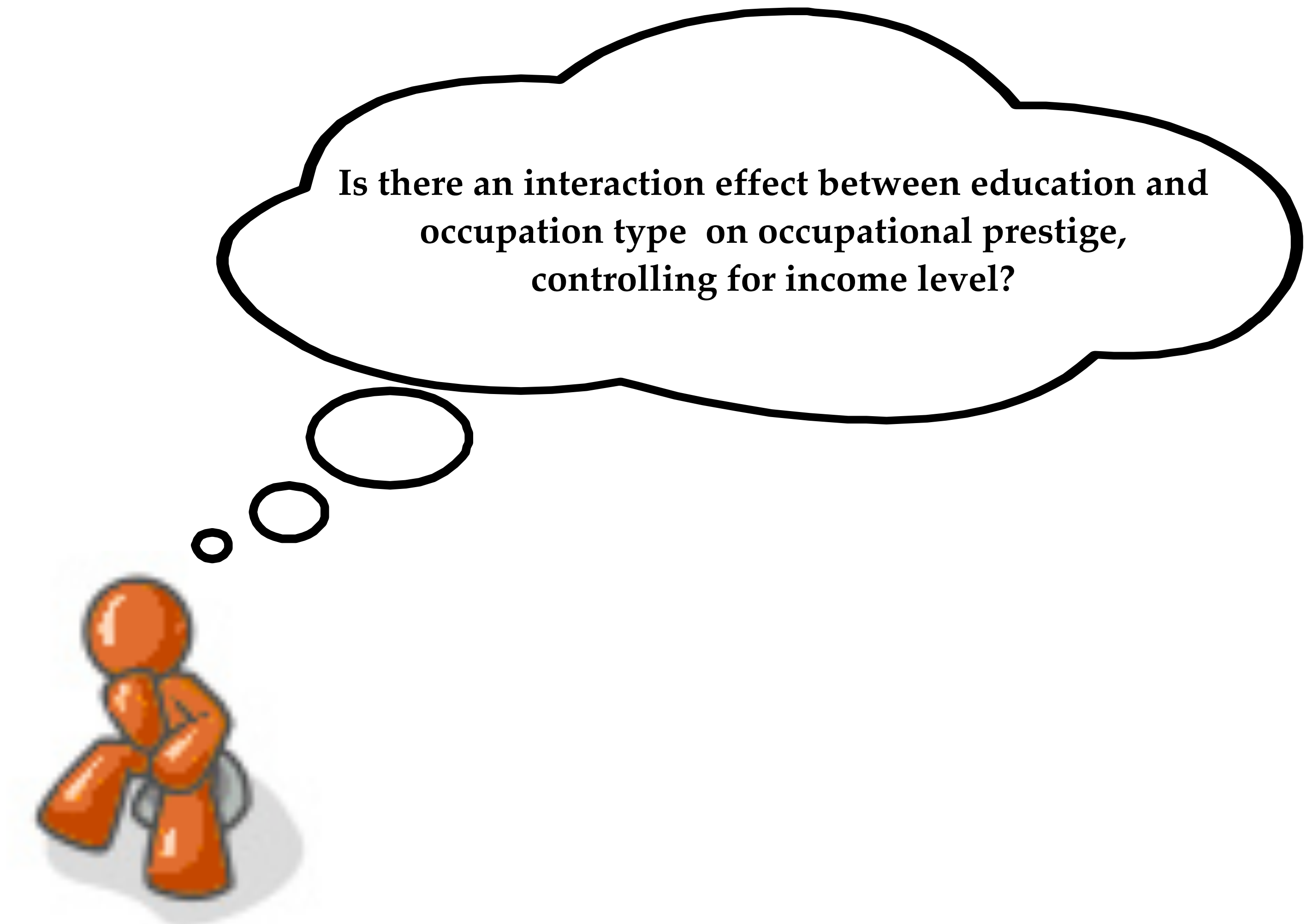
```
lm.bc = lm(prestige ~ 1 + L2income + type + education)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -79.4314    14.1306  -5.621 1.97e-07 ***
L2income      7.2694     1.1900   6.109 2.31e-08 ***
type1        -1.7705     1.8499  -0.957   0.3410
type2         4.9804     1.9416   2.565   0.0119 *
education     3.2845     0.6081   5.401 5.06e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.637 on 93 degrees of freedom
Multiple R-squared:  0.8555,	Adjusted R-squared:  0.8493
F-statistic: 137.6 on 4 and 93 DF,  p-value: < 2.2e-16
```

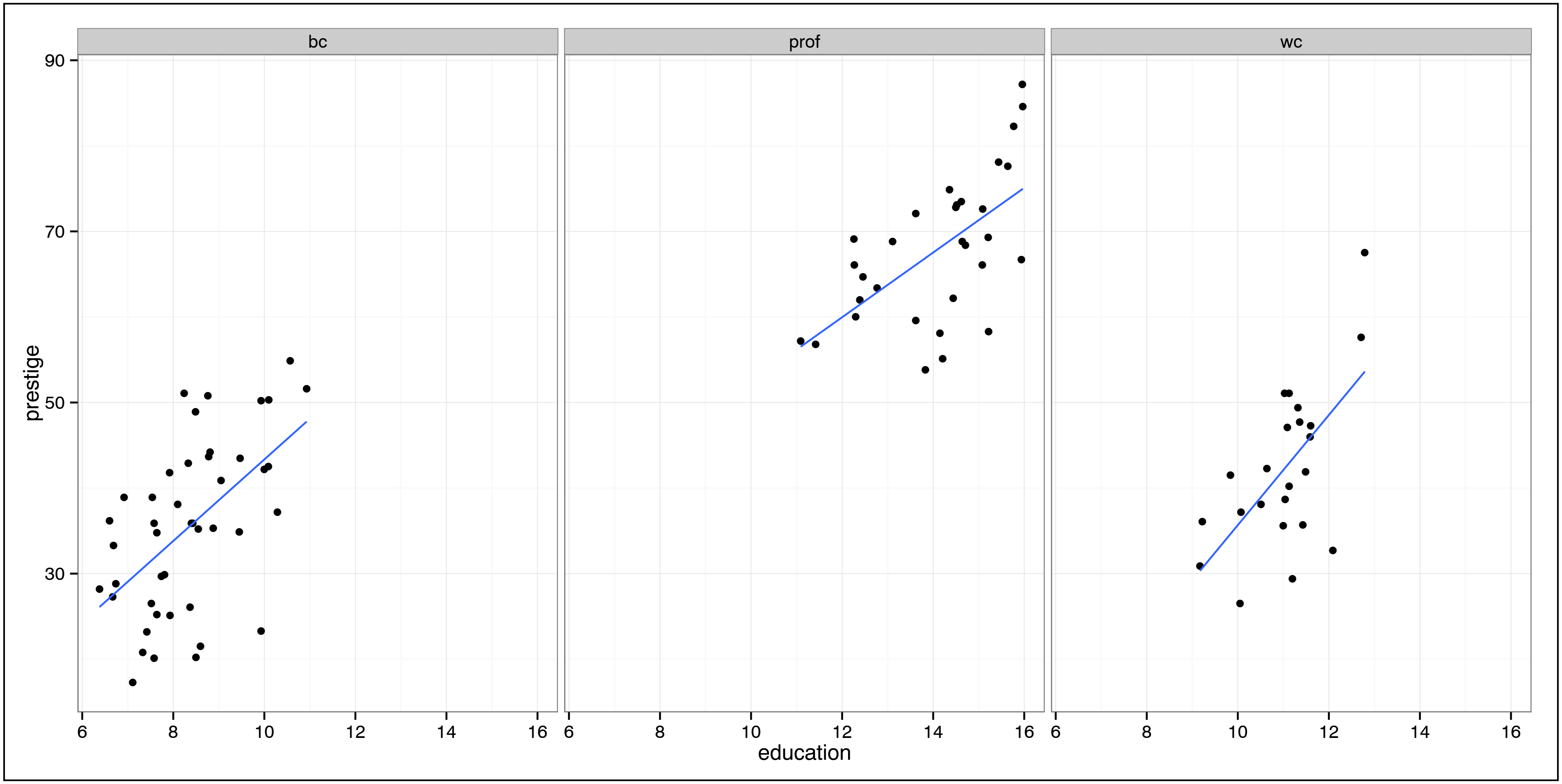
What do the coefficients mean when we use effects coding?

Regardless of the type of contrasts, the plot of the fitted regression lines are the same.




Is there an interaction effect between education and occupation type on occupational prestige, controlling for income level?

Do the data suggest that the effect of education on occupational prestige is the same for all levels of occupation type?



Fit model that includes constituent main-effects and interaction as predictors

Model

```
prestige ~ 1 + L2income + type + education + education:type
```

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        -81.6672    14.3681  -5.684 1.57e-07 ***
L2income             7.4051     1.1858   6.245 1.33e-08 ***
education            3.1407     0.9004   3.488 0.000751 ***
typeprof            15.6176    14.2168   1.099 0.274871
typewc             -30.4466    18.3465  -1.660 0.100451
education:typeprof  -0.5801     1.2211  -0.475 0.635887
education:typewc     2.6675     1.7551   1.520 0.132018
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.585 on 91 degrees of freedom
Multiple R-squared:  0.8608,	Adjusted R-squared:  0.8516
F-statistic: 93.79 on 6 and 91 DF,  p-value: < 2.2e-16
```

If we had fitted the same model using dummy predictors,

Model

```
prestige ~ 1 + L2income + wc + prof + education + education:wc +
     education:prof
```

There are two possible dummy variables for education to interact with, wc or prof.

In context, this means that the effect of education may be different for blue-collar workers, for white-collar workers, or professionals.

This is akin to the idea of mean differences between groups....there are multiple ways that the effect of education might be different for the three occupation types.

Examining the "Omnibus" Interaction Hypothesis

Often the first step when there are multiple possibilities for how an effect may differ is to compare the main-effects model to the interaction model using a ΔF-test.

```
> # Fit main-effects model
> lm.bc.me = lm(prestige ~ L2income + education + type,
    data = Prestige)

> # Fit interaction model
> lm.bc.int = lm(prestige ~ L2income + education + type +
    education:type, data = Prestige)

> # Delta F-test
> anova(lm.bc.me, lm.bc.int)

Analysis of Variance Table

Model 1: prestige ~ L2income + education + type
Model 2: prestige ~ L2income + education + type + education:type
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     93 4096.3
2     91 3945.7  2    150.56 1.7362  0.182
```

If the change in the residual sums-of-squares is statistically reliable, then we examine the summary output, or the plot of the interaction model to see how the interaction plays out.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        -81.6672    14.3681  -5.684 1.57e-07 ***
L2income             7.4051     1.1858   6.245 1.33e-08 ***
education            3.1407     0.9004   3.488 0.000751 ***
typeprof            15.6176    14.2168   1.099 0.274871
typewc             -30.4466    18.3465  -1.660 0.100451
education:typeprof  -0.5801     1.2211  -0.475 0.635887
education:typewc     2.6675     1.7551   1.520 0.132018
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.585 on 91 degrees of freedom
Multiple R-squared:  0.8608, Adjusted R-squared:  0.8516
F-statistic: 93.79 on 6 and 91 DF,  p-value: < 2.2e-16
```

The effect of education for professional occupations seems slightly lower (although not statistically reliable) than the effect of education for blue-collar occupations (the reference group). The effect of education for white-collar occupations seems slightly higher (although not statistically reliable) than the effect of education for blue-collar occupations (the reference group).

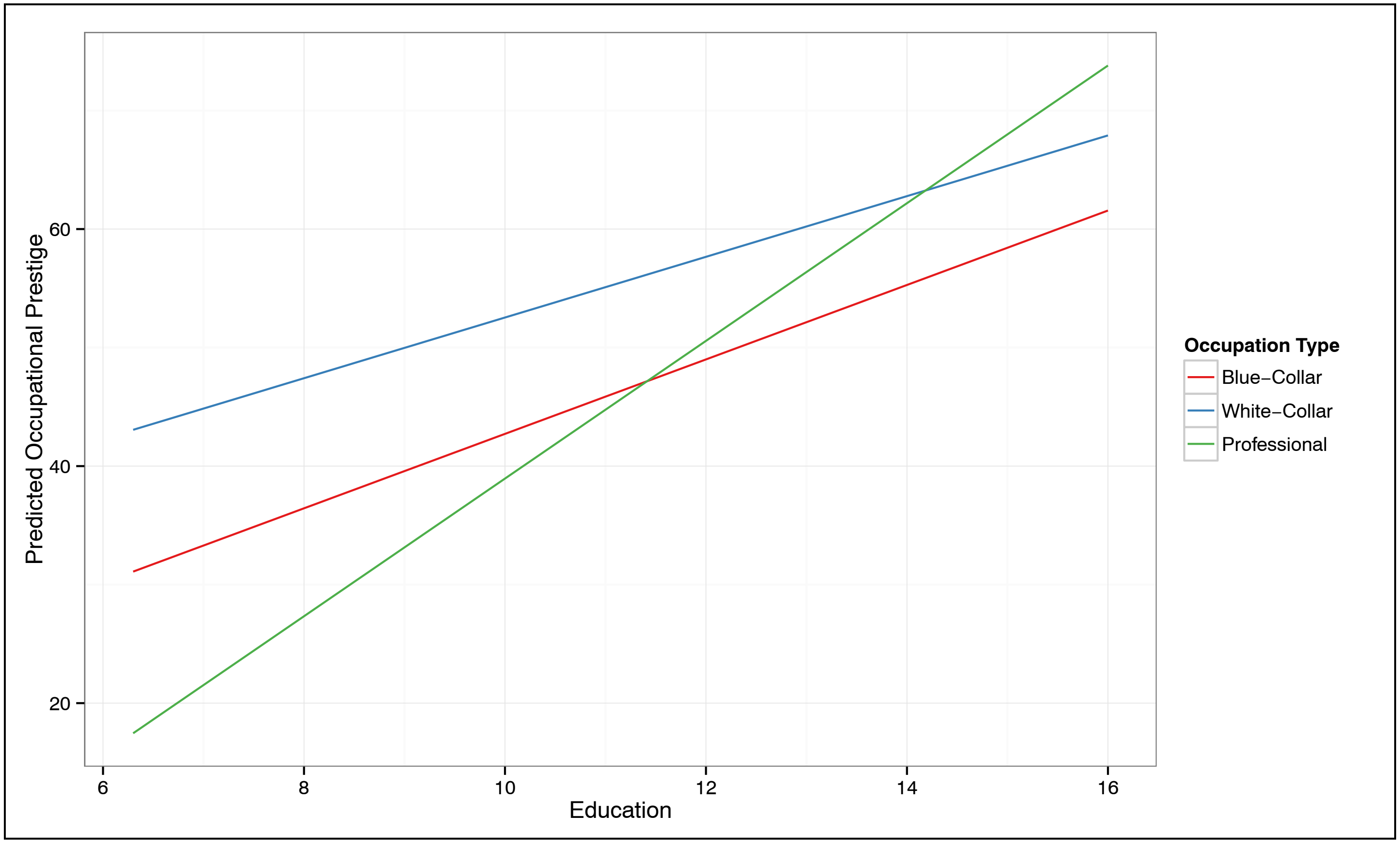
Is the effect of education for white-collar occupations different than the effect of education for professional occupations?

In order to answer this, we would have to change the reference group.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        -112.114     21.477  -5.220 1.12e-06 ***
L2income              7.405      1.186   6.245 1.33e-08 ***
education             5.808      1.524   3.812 0.000251 ***
typebc               30.447     18.346   1.660 0.100451    
typeprof             46.064     20.781   2.217 0.029142 *  
education:typebc     -2.668      1.755  -1.520 0.132018    
education:typeprof   -3.248      1.751  -1.855 0.066819 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.585 on 91 degrees of freedom
Multiple R-squared:  0.8608,	Adjusted R-squared:  0.8516 
F-statistic: 93.79 on 6 and 91 DF,  p-value: < 2.2e-16
```

Note: With multiple comparisons, we should use a *p*-value adjustment.

Plot of the Fitted Values



Although there is an interaction in the sample data, this effect is not statistically reliable…in the population, the lines are parallel!

To Interpret the Effects

If you need to be able to interpret the quantitative `summary()` output, it is best to use the model fitted with dummy variables.

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -81.6672    14.3681  -5.684 1.57e-07 ***
L2income         7.4051     1.1858   6.245 1.33e-08 ***
education        3.1407     0.9004   3.488 0.000751 ***
wc             -30.4466    18.3465  -1.660 0.100451    
prof            15.6176    14.2168   1.099 0.274871    
education:wc     2.6675     1.7551   1.520 0.132018    
education:prof  -0.5801     1.2211  -0.475 0.635887    
---
```

Now we can substitute 0s and 1s in for the dummy variables to get the fitted regression equation for each of the three groups

$$\widehat{\text{Prestige}} = -81.6 + 7.4(\text{L2income}) + 3.1(\text{Education}) - 30.4(\text{wc}) + 15.6(\text{prof}) \\ + 2.7(\text{Education})(\text{wc}) - 0.6(\text{Education})(\text{prof})$$

For blue-collar occupations, `wc=0` and `prof=0`. We will also substitute in the mean value of L2income (12.5) to control that variable out of the model.

$$\widehat{\text{Prestige}} = -81.6 + 7.4(12.5) + 3.1(\text{Education}) - 30.4(0) + 15.6(0) \\ + 2.7(\text{Education})(0) - 0.6(\text{Education})(0)$$

$$\widehat{\text{Prestige}} = -81.6 + 92.5 + 3.1(\text{Education})$$

$$\widehat{\text{Prestige}} = 10.9 + 3.1(\text{Education})$$

The effect of education for blue-collar occupations, controlling for income level, is 3.1. A one-year difference in average education is positively associated with a 3.1-unit difference in average prestige, controlling for income level.

Blue-collar occupations $\widehat{\text{Prestige}} = 10.9 + 3.1(\text{Education})$

White-collar occupations $\widehat{\text{Prestige}} = -19.5 + 5.8(\text{Education})$

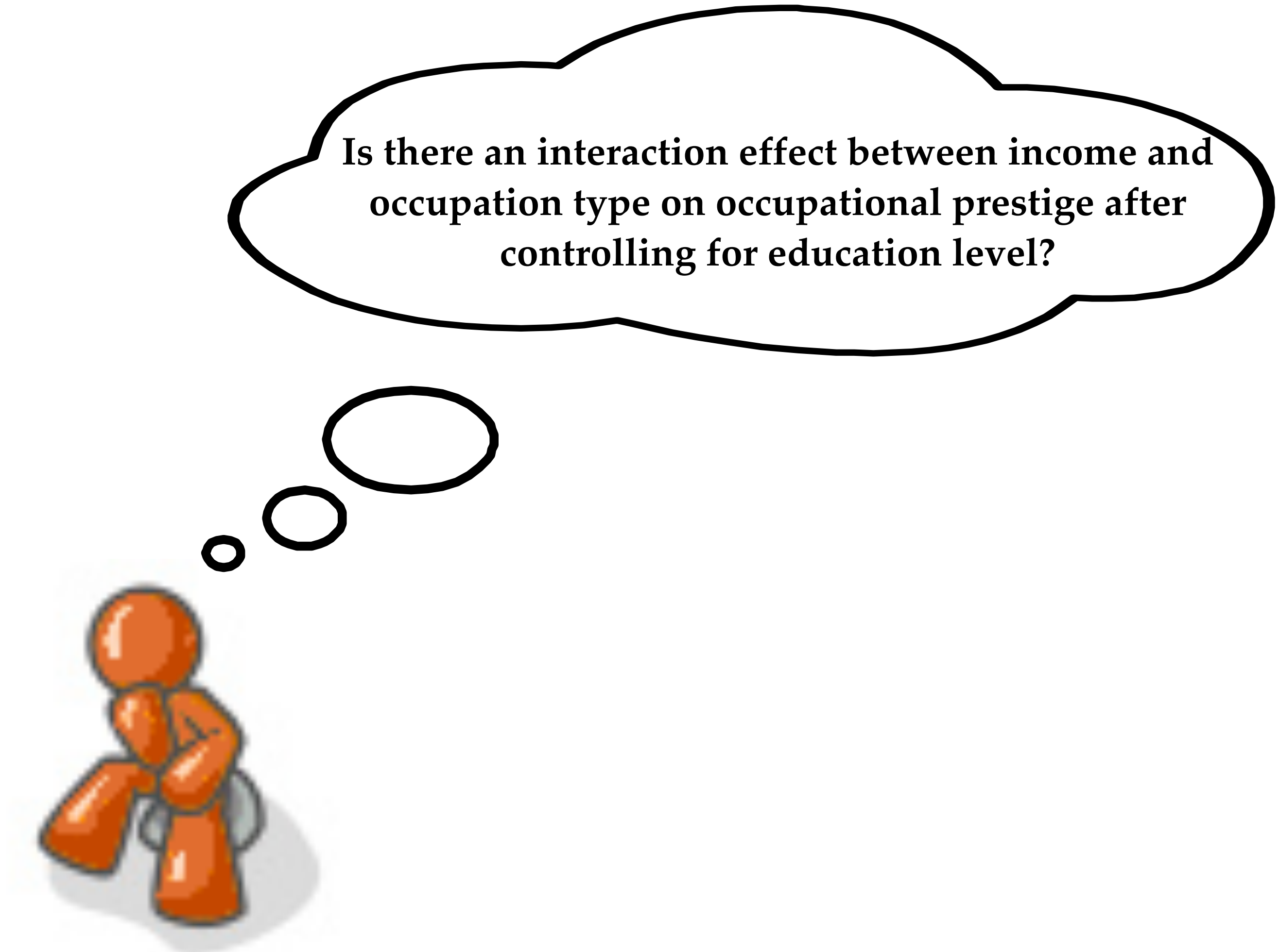
Professional occupations $\widehat{\text{Prestige}} = 26.5 + 2.5(\text{Education})$

Sample Effect

The effect of education on occupational prestige for professional occupations, controlling for income level, is greater than the effect of education on occupational prestige for either blue-collar or white-collar occupations. For these latter two occupation types, the effect of education on occupational prestige is about the same.

Population Effect

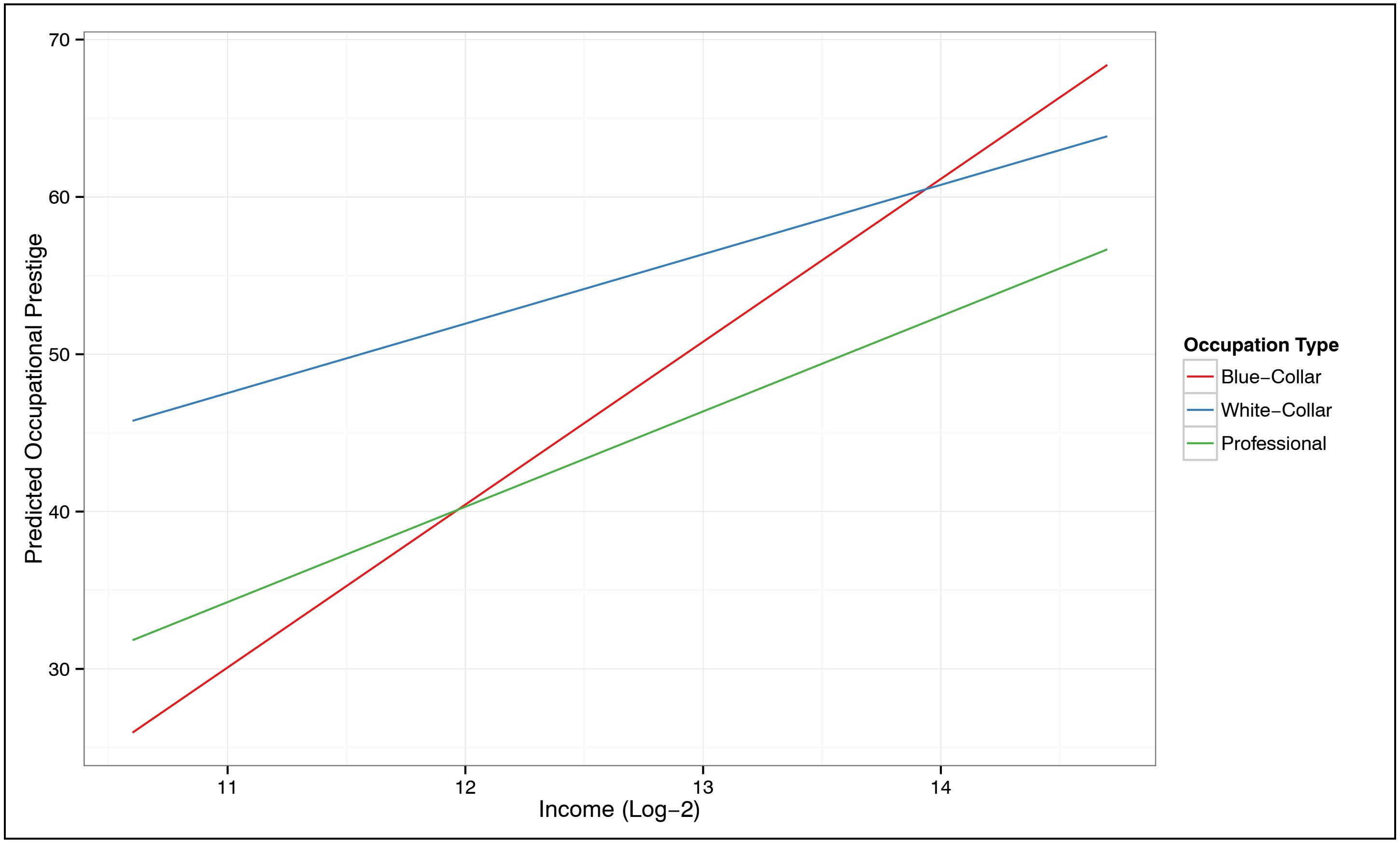
The effect of education on occupational prestige for blue-collar, white-collar, and professional occupations, controlling for income level, is the same.


Is there an interaction effect between income and occupation type on occupational prestige after controlling for education level?

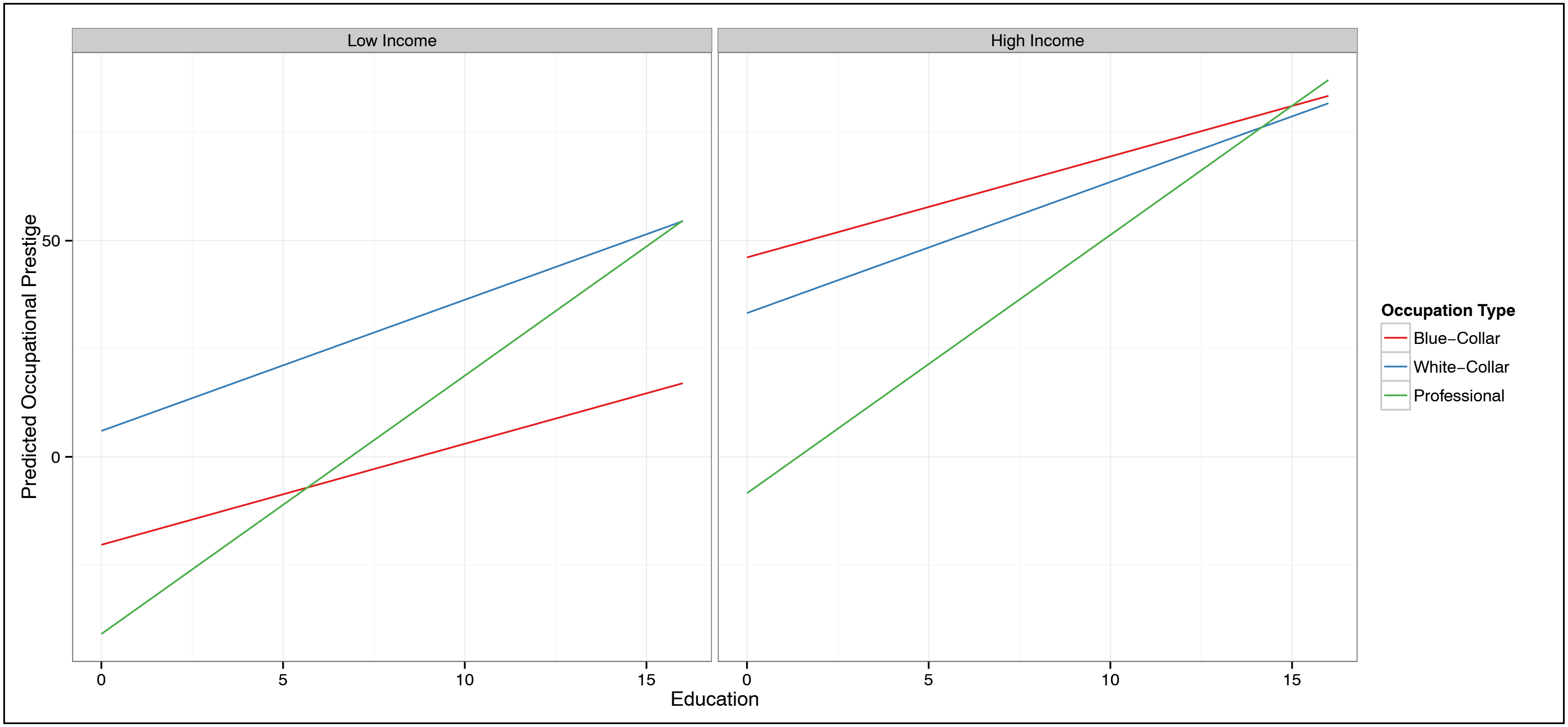
```
> # Fit main-effects model
> lm.bc.me = lm(prestige ~ L2income + education + type,
    data = Prestige)

> # Fit interaction model
> lm.bc.income.int = lm(prestige ~ L2income + education + type +
    L2income:type, data = Prestige)

> # Delta F-test
> anova(lm.bc.me, lm.bc.income.int)

Analysis of Variance Table

Model 1: prestige ~ L2income + education + type
Model 2: prestige ~ L2income + education + type + L2income:type
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     93 4096.3
2     91 3834.2  2    262.13 3.1107 0.04934 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The effect of income on occupational prestige for blue-collar, white-collar, and professional occupations, controlling for income level, is not the same in the population, $p = 0.049$.

Plot of the Fitted Values



Can we include both interactions...occupation type and education AND occupation type and income level



The effect of education on occupational prestige differs by occupation type, controlling for income level. (Seen through the different slopes within a panel)

The effect of income level on occupational prestige differs by occupation type, controlling for education. (Seen through the different slopes across panels)

```
> # Fit main-effects model
> lm.bc.me = lm(prestige ~ L2income + education + type, data = Prestige)

> # Fit interaction model
> lm.bc.both.int = lm(prestige ~ L2income + education + type +
    education:type + L2income:type, data = Prestige)

> # Delta F-test
> anova(lm.bc.me, lm.bc.both.int)

Analysis of Variance Table

Model 1: prestige ~ L2income + education + type
Model 2: prestige ~ L2income + education + type + education:type +
L2income:type
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     93 4096.3
2     89 3655.4  4    440.89 2.6836 0.03646 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

At least one of the interactions is statistically reliable.

```
# Fit interaction model (only education:type)
lm.bc.educ.int = lm(prestige ~ L2income + education + type +
  education:type, data = Prestige)

# Fit interaction model (only L2income:type)
lm.bc.Income.int = lm(prestige ~ L2income + education + type +
  L2income:type, data = Prestige)

# Fit interaction model (both)
lm.bc.both.int = lm(prestige ~ L2income + education + type +
  education:type + L2income:type, data = Prestige)
```

```
# Delta F-test (to test L2income)
> anova(lm.bc.educ.int , lm.bc.both.int)

Analysis of Variance Table

Model 1: prestige ~ L2income + education + type + education:type
Model 2: prestige ~ L2income + education + type + education:type + L2income:type
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     91 3945.7
2     89 3655.4  2    290.33 3.5344 0.03334 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Interaction between L2income and occupation type is important...above and beyond the interaction between education and occupation type

```
# Delta F-test (to test education)
> anova(lm.bc.income.int , lm.bc.both.int)

Analysis of Variance Table

Model 1: prestige ~ L2income + education + type + L2income:type
Model 2: prestige ~ L2income + education + type + education:type + L2income:type
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     91 3834.2
2     89 3655.4  2    178.76 2.1762 0.1195
```

Interaction between education and occupation type is not important...above and beyond the interaction between L2income and occupation type

Examine results

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -89.8047    32.1879  -2.790 0.006447 ** 
L2income              5.4252     2.4599   2.205 0.030002 *  
education             5.9757     1.4944   3.999 0.000131 ***
typebc              -30.2412    37.9788  -0.796 0.427997    
typeprof             54.9189    40.0248   1.372 0.173474    
education:typebc     -3.6400     1.7589  -2.069 0.041404 *  
education:typeprof   -2.9426     1.7422  -1.689 0.094719 .  
L2income:typebc       5.6530     3.0519   1.852 0.067296 .  
L2income:typeprof    -0.8825     3.1041  -0.284 0.776838    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.409 on 89 degrees of freedom
Multiple R-squared:  0.871,	Adjusted R-squared:  0.8595 
F-statistic: 75.15 on 8 and 89 DF,  p-value: < 2.2e-16
```

These are different than the results of the ΔF-test.