More Interaction Models

2018-11-26

Preparation

In this set of notes, you will continue to learn about interaction models. To do so, we will examine the question of whether there is a differential effect of beauty by age on course evaluation scores. The data we will use in this set of notes is collected from student evaluations of instructors' beauty and teaching quality for several courses at the University of Texas. The teaching evaluations were conducted at the end of the semester, and the beauty judgments were made later, by six students who had not attended the classes and were not aware of the course evaluations. The variables are:

- `prof_id`: Professor ID number
- `avg_eval`: Average course rating
- `num_courses`: Number of courses for which the professor has evaluations
- `num_students`: Number of students enrolled in the professor's courses
- `perc_evaluating`: Average percentage of enrolled students who completed an evaluation
- `beauty`: Measure of the professor's beauty composed of the average score on six standardized beauty ratings
- `tenured`: Is the professor tenured? (0 = non-tenured; 1 = tenured)
- `native_english`: Is the professor a native English speaker? (0 = non-native English speaker; 1 = native English speaker)
- `age`: Professor's age (in years)
- `female`: Is the professor female? (0 = male; 1 = female)

These source of these data is: Hamermesh, D. S. & Parker, A. M. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*, 369–376. The data were made available by: Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

```r
# Load libraries
library(broom)
library(corrr)
library(dotwhisker)
library(dplyr)
library(ggplot2)
library(readr)
library(sm)
library(tidyr)

# Read in data
evals = read_csv(file = "~/Documents/github/epsy-8251/data/evaluations.csv")
```

Interaction between Age and Beauty

Typically, barring support for the interaction effect from theoretical/substantive findings, we would explore the sample data for empirical evidence of the interaction (generally via plots of the data). To explore an interaction effect between two quantitative variables poses some unique challenges.

To understand those challeneges consider how we explored the interaction between sex and beauty. We created a plot of the effect of beauty on average course evaluation score for males and females, and asked whether the regression line for males and females were parallel. In other words, we need to examine the relationship between $X1$ and Y for different levels of $X2$.

If we are examining an interaction between age and beauty on course evaluation scores, we need to again examine the effect of beauty on average course evaluation score at different levels of age. But, since age is continuous, we have to choose the levels of age; they aren't pre-specified like males and females. In general, researchers would probably choose a high and low (or high, medium, and low) value of age.

So, we are going to examine the relationship between beauty and course evaluation scores for a low, medium, and high age. Empirically, we can choose these values based on the `summary()` output.

```r
summary(evals)
```

```
    prof_id         avg_eval      num_courses      num_students   
 Min.   : 1.00   Min.   :2.300   Min.   : 1.000   Min.   :  12.0  
 1st Qu.:24.25   1st Qu.:3.646   1st Qu.: 3.000   1st Qu.:  83.0  
 Median :47.50   Median :3.937   Median : 4.000   Median : 169.0  
 Mean   :47.50   Mean   :3.906   Mean   : 4.926   Mean   : 271.8  
 3rd Qu.:70.75   3rd Qu.:4.219   3rd Qu.: 7.000   3rd Qu.: 290.8  
 Max.   :94.00   Max.   :4.787   Max.   :13.000   Max.   :2798.0  
 perc_evaluating     beauty             tenured       native_english  
 Min.   :33.35   Min.   :-1.5388430   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:66.83   1st Qu.:-0.6690352   1st Qu.:0.0000   1st Qu.:1.0000  
 Median :75.48   Median :-0.1281433   Median :1.0000   Median :1.0000  
 Mean   :73.57   Mean   : 0.0000001   Mean   :0.5851   Mean   :0.9255  
 3rd Qu.:82.14   3rd Qu.: 0.6157626   3rd Qu.:1.0000   3rd Qu.:1.0000  
 Max.   :94.48   Max.   : 1.8816740   Max.   :1.0000   Max.   :1.0000  
      age            female      
 Min.   :29.00   Min.   :0.0000  
 1st Qu.:39.00   1st Qu.:0.0000  
 Median :47.00   Median :0.0000  
 Mean   :47.55   Mean   :0.4255  
 3rd Qu.:56.00   3rd Qu.:1.0000  
 Max.   :73.00   Max.   :1.0000  
```

Here we might choose ages of 40 (low), 50 (medium), and 60 (high). after choosing particular values of age, we run into another challenge. That is, there are typically only a few observations that have those ages.

```r
evals_2 = evals %>% filter(age == 40 | age == 50 | age == 60)
nrow(evals_2)
```

```
[1] 7
```

There are only 7 cases (out of 94) that are at those three ages. Ideally, we want to use the entire data set to examine effects, otherwise, we might see a spurious relationship.

Cutting the Age Variable into Categories

For plotting purposes, what is usually done is to "cut" the continuous predictor into discrete categories (e.g., low ages, medium ages, and high ages). This is ONLY done for plotting. When we fit the actual interaction, we use the continuous predictor.

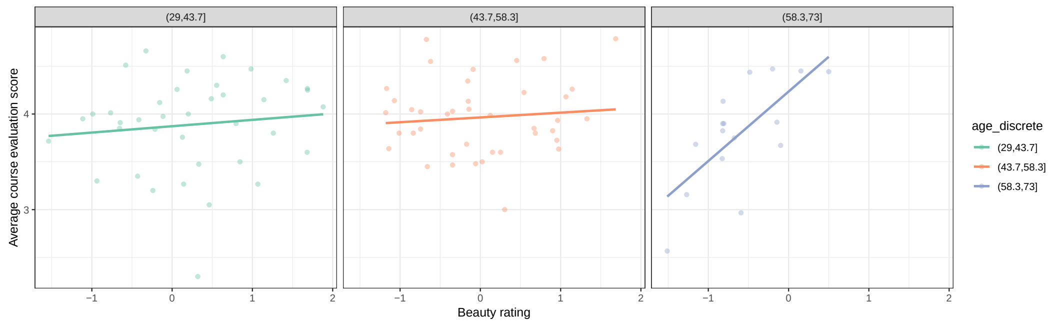
There are several R functions to dicretize a continuous variable. We will use the `cut()` function. This function takes the name of the variable you want to cut, and an argument `break=` which specifies the number of levels you want the variable cut into. (Note: you can also specify the cutpoints in the `breaks=` argument; see `cut()`'s help menu.)

```
evals = evals %>%
  mutate(
    age_discrete = cut(age, breaks = 3)
  )

summary(evals)
```

```
    prof_id         avg_eval      num_courses      num_students
 Min.   : 1.00   Min.   :2.300   Min.   : 1.000   Min.   :  12.0
 1st Qu.:24.25   1st Qu.:3.646   1st Qu.: 3.000   1st Qu.:  83.0
 Median :47.50   Median :3.937   Median : 4.000   Median : 169.0
 Mean   :47.50   Mean   :3.906   Mean   : 4.926   Mean   : 271.8
 3rd Qu.:70.75   3rd Qu.:4.219   3rd Qu.: 7.000   3rd Qu.: 290.8
 Max.   :94.00   Max.   :4.787   Max.   :13.000   Max.   :2798.0
 perc_evaluating     beauty               tenured       native_english
 Min.   :33.35   Min.   :-1.5388430   Min.   :0.0000   Min.   :0.0000
 1st Qu.:66.83   1st Qu.:-0.6690352   1st Qu.:0.0000   1st Qu.:1.0000
 Median :75.48   Median :-0.1281433   Median :1.0000   Median :1.0000
 Mean   :73.57   Mean   : 0.0000001   Mean   :0.5851   Mean   :0.9255
 3rd Qu.:82.14   3rd Qu.: 0.6157626   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :94.48   Max.   : 1.8816740   Max.   :1.0000   Max.   :1.0000
      age            female          age_discrete
 Min.   :29.00   Min.   :0.0000   (29,43.7]   :38
 1st Qu.:39.00   1st Qu.:0.0000   (43.7,58.3]:40
 Median :47.00   Median :0.0000   (58.3,73]   :16
 Mean   :47.55   Mean   :0.4255
 3rd Qu.:56.00   3rd Qu.:1.0000
 Max.   :73.00   Max.   :1.0000
```

Now we have cut age into three levels: low ages (43.7 or younger), medium ages (older than 43.7, and younger than or equal to 58.3), and high ages (older than 58.3). We can use our new variable to now examine the potential interaction.

```
ggplot(data = evals, aes(x = beauty, y = avg_eval, color = age_discrete)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm", se = FALSE) +
    theme_bw() +
    xlab("Beauty rating") +
    ylab("Average course evaluation score") +
  scale_color_brewer(palette = "Set2") +
  facet_wrap(~age_discrete)
```



Fit the Interaction Model

To fit the interaction model, use the consituent main effects and the interaction term to predict course evaluation scores. VERY IMPORTANT—Use the original quantitative age predictor, not the discretized variable in the model. We will also use the colon (`:`) notation to fit the model. The colon implicitly creates the product term and includes it in the model.

```
# Fit model
lm.1 = lm(avg_eval ~ 1 + beauty + age + beauty:age, data = evals)

# Model-level output
glance(lm.1)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
*     <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>  <dbl> <dbl> <dbl>
1    0.0616        0.0303 0.454      1.97   0.124     4  -57.1  124.  137.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```
# Coefficient-level output
tidy(lm.1)
```

```
# A tibble: 4 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  3.77      0.239      15.8   1.21e-27
2 beauty      -0.249     0.266      -0.937 3.51e- 1
3 age          0.00344   0.00497     0.692 4.91e- 1
4 beauty:age   0.00807   0.00569     1.42  1.59e- 1
```

The interaction effect is not statistically significant, indicating there is likely not an interaction between age and beauty on course evaluation scores in the population.

Plot of the Interaction Model

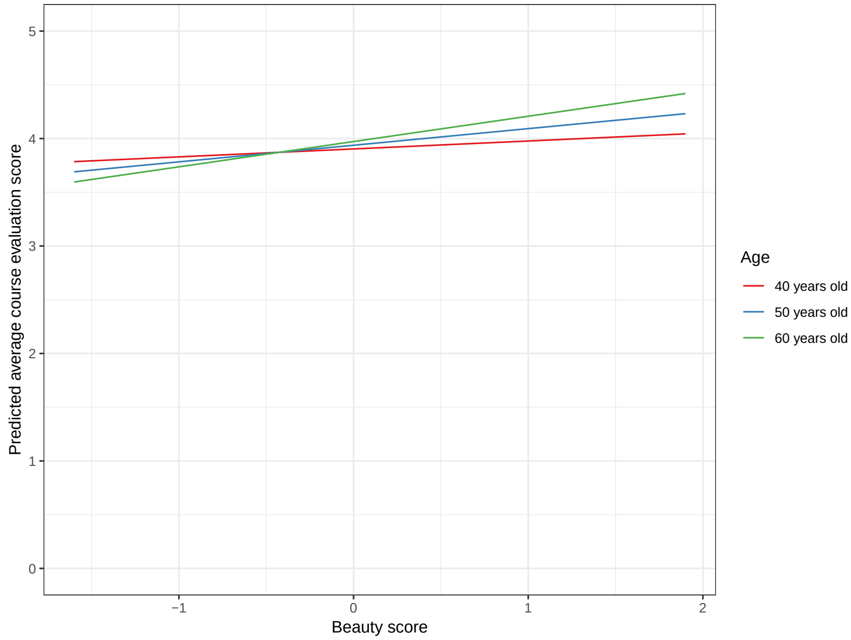
Although, in practice, we would not interpret a non-significant interaction, we will continue as if it were significant for a pedagogical exercise.

When we plot the model results, we will choose specific values for age and include them in our `crossing()` function along with a sequence of beauty ratings (and sex).

```r
# Create new data set with main effects
plot_data = crossing(
  beauty = seq(from = -1.6, to = 1.9, by = 0.1),
  age = c(40, 50, 60)
)

# Use fitted model to compute fitted values for the data
plot_data = plot_data %>%
  mutate(
    yhat = predict(lm.1, newdata = plot_data)
  )

head(plot_data)
```

```
# A tibble: 6 x 3
  beauty   age  yhat
   <dbl> <dbl> <dbl>
1   -1.6    40  3.79
2   -1.6    50  3.69
3   -1.6    60  3.60
4   -1.5    40  3.79
5   -1.5    50  3.71
6   -1.5    60  3.62
```

```r
# Plot the fitted model
ggplot(data = plot_data, aes(x = beauty, y = yhat, color = factor(age))) +
  geom_line() +
  theme_bw() +
  xlab("Beauty score") +
  ylab("Predicted average course evaluation score") +
  scale_color_brewer(
    name = "Age",
    palette = "Set1",
    labels = c("40 years old", "50 years old", "60 years old")
    ) +
  ylim(0, 5)
```



Based on the plot, we can see there is a *disordinal interaction* between age and beauty (the lines cross in our plot). The effect of beauty on average course evaluation score varies across the three ages. There largest effect of beauty on average course evaluation score is for older professors (highest slope), while the effect of beauty for younger professors is smaller (lower slopes).

Similarly, the effect of age on average course evaluation score varies across beauty rating. For professors who were percieved as having a less than average beauty rating (< 0) younger professors tend to receive higher course evaluation ratings than older professors. However, this trend reverses itself for professors who were percieved as having a higher than average beauty rating. For those professors, older professors tend to be given higher course evaluation ratings than younger professors.

Compute the Fitted Equations

We can also compute the fitted equations for professors whose age varies. To do this, substitute in the age values in the fitted regression equation. For young, male professors (age = 40),

$$\begin{aligned}
\hat{\text{course eval}}_i &= 3.77 - 0.25(\text{beauty}_i) + 0.003(\text{age}_i) + 0.008(\text{beauty}_i)(\text{age}_i) \\
&= 3.77 - 0.25(\text{beauty}_i) + 0.003(40) + 0.008(\text{beauty}_i)(40) \\
&= 3.77 - 0.25(\text{beauty}_i) + 0.12 + 0.32(\text{beauty}_i) \\
&= 3.89 + 0.07(\text{beauty}_i)
\end{aligned}$$

For 50 year old professors,

$$\begin{aligned}
\hat{\text{course eval}}_i &= 3.77 - 0.25(\text{beauty}_i) + 0.003(\text{age}_i) + 0.008(\text{beauty}_i)(\text{age}_i) \\
&= 3.77 - 0.25(\text{beauty}_i) + 0.003(50) + 0.008(\text{beauty}_i)(50) \\
&= 3.77 - 0.25(\text{beauty}_i) + 0.15 + 0.40(\text{beauty}_i) \\
&= 3.92 + 0.15(\text{beauty}_i)
\end{aligned}$$

And, finally, for 60 year old professors,

$$
\begin{aligned}
\hat{\text{course eval}}_i &= 3.77 - 0.25(\text{beauty}_i) + 0.003(\text{age}_i) + 0.008(\text{beauty}_i)(\text{age}_i) \\
&= 3.77 - 0.25(\text{beauty}_i) + 0.003(60) + 0.008(\text{beauty}_i)(60) \\
&= 3.77 - 0.25(\text{beauty}_i) + 0.18 + 0.48(\text{beauty}_i) \\
&= 3.95 + 0.23(\text{beauty}_i)
\end{aligned}
$$

This tells us the same thing we saw in the plot, but helps us see the difference in slopes numerically.

Interpreting the Individual Effects from the tidy() Output

```r
tidy(lm.1)
```

```
# A tibble: 4 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  3.77      0.239      15.8   1.21e-27
2 beauty      -0.249     0.266      -0.937 3.51e- 1
3 age          0.00344   0.00497     0.692 4.91e- 1
4 beauty:age   0.00807   0.00569     1.42  1.59e- 1
```

In practice, use the plot of the results to interpret interaction effects. In simple models, we can actually interpret the coefficients more directly. To do this, write out the fitted equations for professors who differ in age by 1 year. We will do this for professors who are 0 years old and professors who are 1 year old. (Do the substitution yourself to verify these equations.)

$$
\begin{aligned}
\textbf{0 year olds}: &\quad \hat{\text{course eval}}_i = 3.77 - 0.25(\text{beauty}_i) \\
\textbf{1 year olds}: &\quad \hat{\text{course eval}}_i = [3.77 + 0.003] + [-0.25 + 0.008]\,(\text{beauty}_i)
\end{aligned}
$$

- The intercept (3.77) is the average course evaluation score for professors with a 0 beauty rating who are 0 years old (extrapolation).
- The coefficient associated with beauty (-0.25) is the effect of beauty for professors who are 0 years old.
- The coefficient associated with age (0.003) is the difference in average course evaluation between professors with beauty rating = 0 whose age differs by one year.
- The coefficient associated with the interaction term (0.008) is the difference in slopes (effect of beauty on course evaluation) between professors whose age differs by one year.

Note that although we can interpret the coefficients directly, the plot is typically more informative and far less complicated to make sense of.

Adding Covariates

Is there an interaction between age and beauty on average course evaluations after we control for gender differences?

```
# Fit model
lm.2 = lm(avg_eval ~ 1 + beauty + age + female + beauty:age, data = evals)

# Model-level output
glance(lm.2)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
*     <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>  <dbl> <dbl> <dbl>
1     0.130        0.0910 0.440      3.33  0.0137     5  -53.6  119.  134.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```
# Coefficient-level output
tidy(lm.2)
```

```
# A tibble: 5 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  3.97      0.244      16.3   1.94e-28
2 beauty      -0.304     0.258      -1.18  2.42e- 1
3 age          0.00147   0.00487     0.302 7.64e- 1
4 female      -0.252     0.0951     -2.65  9.57e- 3
5 beauty:age   0.00965   0.00554     1.74  8.49e- 2
```

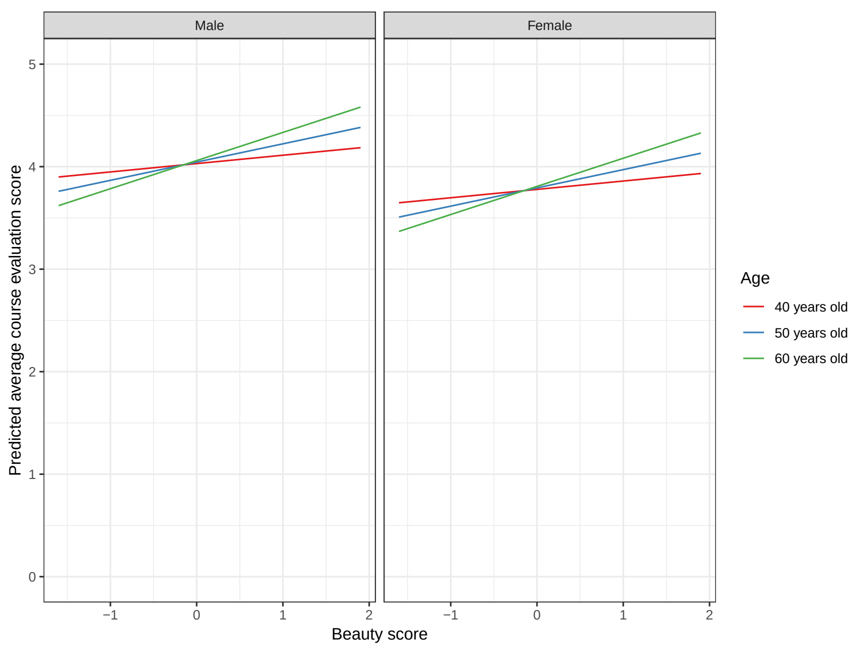
Yes. The interaction term is statsitically significant after we include sex of the professor in the model. To understand this model, plot it.

```
# Create new data set with main effects
plot_data = crossing(
  beauty = seq(from = -1.6, to = 1.9, by = 0.1),
  age = c(40, 50, 60),
  female = c(0, 1)
)

# Use fitted model to compute fitted values for the data
plot_data = plot_data %>%
  mutate(
    yhat = predict(lm.2, newdata = plot_data),
    age = factor(age),
    female = factor(female, levels = c(0, 1), labels = c("Male", "Female"))
  )

head(plot_data)
```

```
# A tibble: 6 x 4
  beauty age   female  yhat
   <dbl> <fct> <fct>  <dbl>
```

```
1   -1.6 40    Male    3.90
2   -1.6 40    Female  3.65
3   -1.6 50    Male    3.76
4   -1.6 50    Female  3.51
5   -1.6 60    Male    3.62
6   -1.6 60    Female  3.37
```

```r
# Plot the fitted model
ggplot(data = plot_data, aes(x = beauty, y = yhat, color = age)) +
  geom_line() +
  theme_bw() +
  xlab("Beauty score") +
  ylab("Predicted average course evaluation score") +
  scale_color_brewer(
    name = "Age",
    palette = "Set1",
    labels = c("40 years old", "50 years old", "60 years old")
    ) +
  ylim(0, 5) +
  facet_wrap(~female)
```



The plot shows that the effect of beauty on average course evaluation varies for professors with different ages. In general, there is a greater effect of beauty for older professors. This interaction is THE SAME for males and females, although on average, females have lower course evaluations than males for the same beauty rating and age.

As an exercise, compute the fitted equations for males who are 0 and 1 years old (one year difference) and for females who are 0 and 1 years old (four total equations). Then go back to the `tidy()` output and use your equations to help interpret the regression coefficients.

Higher Order Interactions

Interactions between two predictors (e.g., age and beauty) are referred to as *first order* interactions. In the previous section, the model we fitted included a main-effect of gender and a first order interaction between age and beauty. The main-effect of sex in this model suggested that the first order interaction between beauty and age was THE SAME for males and females.

We could also fit a model that posits that the first order interaction between beauty and age IS DIFFERENT for males and females. This is technically an interaction between sex and the first order interaction between beauty and age. It is an interaction of an interaction. This is called a *second order* interaction.

To fit such a model, we would need to include the second order interaction between gender, beauty and age; the product of the three main effects. Since we have an interaction, we need to include all constituent main effects AND since it is a higher order interaction, we need to include all constituent lower order interactions; in this case all constituent first order interactions. As such the predictors would include:

- **Main-Effects:** `beauty` and `age` and `female`
- **First Order Interactions:** `beauty:age` and `beauty:female` and `age:female`
- **Second Order Interaction:** `beauty:age:female`

We fit the model below.

```r
# Fit model
lm.3 = lm(avg_eval ~ 1 + beauty + age + female +
            beauty:age + beauty:female + female:age +
            beauty:age:female, data = evals)

# Model-level output
glance(lm.3)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
*     <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>  <dbl> <dbl> <dbl>
1     0.210         0.146 0.426      3.27 0.00397     8  -49.0  116.  139.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```r
# Coefficient-level output
tidy(lm.3)
```

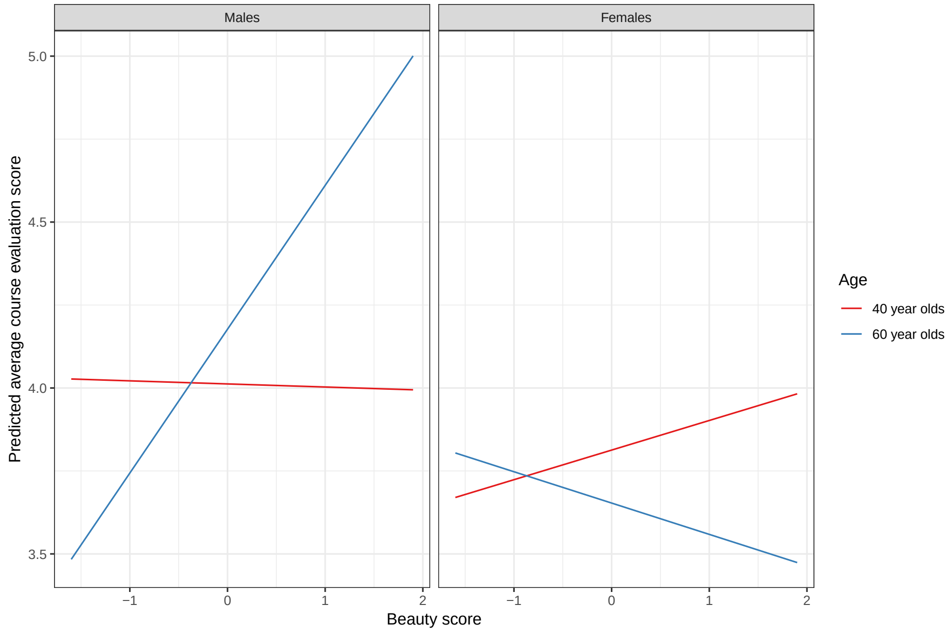
```
# A tibble: 8 x 5
  term              estimate std.error statistic  p.value
  <chr>                <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)        3.68      0.296      12.4   6.28e-21
2 beauty            -0.895     0.376      -2.38  1.95e- 2
3 age                0.00824   0.00600     1.37  1.73e- 1
4 female             0.450     0.465       0.967 3.36e- 1
5 beauty:age         0.0221    0.00746     2.97  3.90e- 3
6 beauty:female      1.35      0.553       2.44  1.66e- 2
7 age:female        -0.0162    0.00983    -1.65  1.02e- 1
8 beauty:age:female -0.0313    0.0122     -2.57  1.20e- 2
```

The second order interaction term is statistically significant. To interpret this, plot the model results.

```r
# Create new data set with main effects
plot_data = crossing(
  beauty = seq(from = -1.6, to = 1.9, by = 0.1),
  age = c(40, 60),
  female = c(0, 1)
)

# Use fitted model to compute fitted values for the data
plot_data = plot_data %>%
  mutate(
    yhat = predict(lm.3, newdata = plot_data)
    )

# Convert female and age into factors for better plotting
plot_data = plot_data %>%
  mutate(
    Sex = factor(female, levels = c(0, 1), labels = c("Males", "Females")),
    Age = factor(age, levels = c(40, 60), labels = c("40 year olds", "60 year olds"))
  )

head(plot_data)
```

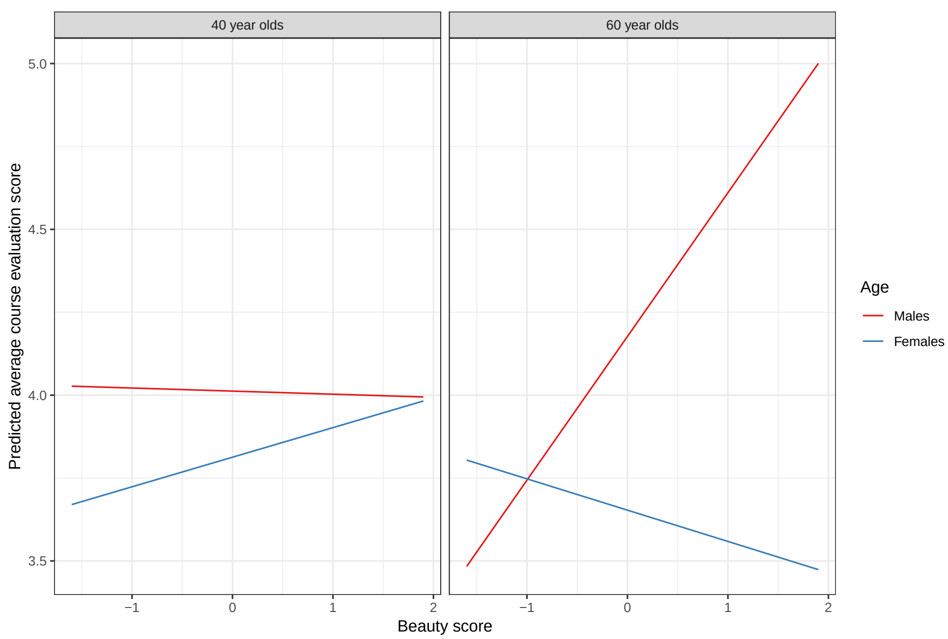
```
# A tibble: 6 x 6
  beauty   age female  yhat Sex     Age
   <dbl> <dbl>  <dbl> <dbl> <fct>   <fct>
1   -1.6    40      0  4.03 Males   40 year olds
2   -1.6    40      1  3.67 Females 40 year olds
3   -1.6    60      0  3.48 Males   60 year olds
4   -1.6    60      1  3.80 Females 60 year olds
5   -1.5    40      0  4.03 Males   40 year olds
6   -1.5    40      1  3.68 Females 40 year olds
```

```r
# Plot the fitted model
ggplot(data = plot_data, aes(x = beauty, y = yhat, color = Age)) +
  geom_line() +
  theme_bw() +
  xlab("Beauty score") +
  ylab("Predicted average course evaluation score") +
  scale_color_brewer(name = "Age", palette = "Set1") +
  facet_wrap(~Sex)
```



- The plots show that the interaction between age and beauty on course evaluation scores DIFFERS for males and females.
- This also suggests that the interaction between beauty and sex on course evaluation scores DIFFERS for professors of different ages

```
ggplot(data = plot_data, aes(x = beauty, y = yhat, color = Sex)) +
  geom_line() +
  theme_bw() +
  xlab("Beauty score") +
  ylab("Predicted average course evaluation score") +
  scale_color_brewer(name = "Age", palette = "Set1") +
  facet_wrap(~Age)
```



Lastly, it also implies that the interaction between sex and age on course evaluation scores DIFFERS for professors with different beauty ratings (not shown; left as an exercise for the reader).

Some Advice for Fitting Interaction Models

In general, only fit interaction terms that include focal predictors. Do not fit interaction terms that are composed of all control predictors. This has the implication that if you do not have a focal predictor (i.e., the analysis is purely exploratory) you should probably not fit interaction terms.

A second piece of advice is that unless there is specific theoretical reason to fit higher order interactions with your focal predictors, avoid them. This also is good advice for first order interaction terms as well.