Business Statistics 41000

MBA Grades

Mladen Kolar

Interactions with continuous variables

Example:connection between college & MBA grades.
A model to predict Booth GPA from college GPA could be

$$\mathrm{GPA}^{\mathrm{MBA}} = \beta_0 + \beta_1 \mathrm{GPA}^{\mathrm{Bach}} + \varepsilon.$$

```r
grades <- read.csv("grades.csv")
summary(grades)
```

```
##       Age          BachGPA            GMAT           MBAGPA     
##  Min.   :24   Min.   :2.30   Min.   :570   Min.   :2.50  
##  1st Qu.:26   1st Qu.:3.10   1st Qu.:640   1st Qu.:3.38  
##  Median :27   Median :3.40   Median :680   Median :3.50  
##  Mean   :28   Mean   :3.36   Mean   :677   Mean   :3.47  
##  3rd Qu.:29   3rd Qu.:3.56   3rd Qu.:710   3rd Qu.:3.70  
##  Max.   :42   Max.   :3.98   Max.   :790   Max.   :4.00  
```

```r
summary(lm(MBAGPA ~ BachGPA, data=grades))
```

```
##
## Call:
## lm(formula = MBAGPA ~ BachGPA, data = grades)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8727 -0.1105  0.0276  0.1491  0.7009
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.5899     0.3121    8.30  1.2e-11 ***
## BachGPA       0.2627     0.0924    2.84   0.0061 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.262 on 62 degrees of freedom
## Multiple R-squared:  0.115,  Adjusted R-squared:  0.101
## F-statistic: 8.08 on 1 and 62 DF,  p-value: 0.00607
```

- For every 1 point increase in college GPA, your expected GPA at Booth increases by about 0.26 points.

However, this model assumes that the marginal effect of College GPA is the same for any age.

- But I'd guess that how you did in college has less effect on your MBA GPA as you get older (farther from college).

We can account for this intuition with an interaction term:

$$\mathrm{GPA}^{\mathrm{MBA}} = \beta_0 + \beta_1 \mathrm{GPA}^{\mathrm{Bach}} + \beta_2(\mathrm{Age} \times \mathrm{GPA}^{\mathrm{Bach}}) + \varepsilon$$

Now, the college effect is

$$\frac{\partial E[\mathrm{GPA}^{\mathrm{MBA}} \mid \mathrm{GPA}^{\mathrm{Bach}}, \mathrm{Age}]}{\partial \mathrm{GPA}^{\mathrm{Bach}}} = \beta_1 + \beta_2 \mathrm{Age}.$$

$\Rightarrow$ Depends on Age!

Fitting interactions in `R` is easy:

`lm(Y ~ X1*X2)` fits $E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$.

Here, we want the interaction but do not want to include the main effect of age (should age matter individually?).

```
summary(lm(MBAGPA ~ BachGPA*Age - Age, data=grades))
```

```
##
## Call:
## lm(formula = MBAGPA ~ BachGPA * Age - Age, data = grades)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7006 -0.1148  0.0003  0.1465  0.6832
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.82049    0.29693    9.50  1.2e-13 ***
## BachGPA      0.45575    0.10303    4.42  4.1e-05 ***
## BachGPA:Age -0.00938    0.00279   -3.37   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.243 on 61 degrees of freedom
## Multiple R-squared:  0.254,  Adjusted R-squared:  0.229
## F-statistic: 10.4 on 2 and 61 DF,  p-value: 0.000132
```

Without the interaction term

- Marginal effect of College GPA is $b_1 = 0.26$.

With the interaction term:

- Marginal effect is $b_1 + b_2 \mathrm{Age} = 0.46 - 0.0094 \mathrm{Age}$.

Age	Marginal Effect
25	0.22
30	0.17
35	0.13
40	0.08

With the main effect for age

```
summary(lm(MBAGPA ~ BachGPA*Age))
```

```
##
## Call:
## lm(formula = MBAGPA ~ BachGPA * Age)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.781 -0.128  0.020  0.146  0.674
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2796     2.5926   -0.11    0.914
## BachGPA       1.3694     0.7660    1.79    0.079 .
## Age           0.1097     0.0912    1.20    0.233
## BachGPA:Age  -0.0418     0.0271   -1.54    0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.242 on 60 degrees of freedom
## Multiple R-squared:  0.271,  Adjusted R-squared:  0.235
## F-statistic: 7.45 on 3 and 60 DF,  p-value: 0.000254
```