Interaction Effects

In this tutorial, we expand our linear regression framework to include **interaction effects**. An interaction effect is when the *combination* of two variables has a *different effect* on the outcome variable than just the sum of the impact from each variable in isolation.

We will illustrate the meaning and use of interaction effects with an example

1. Example: Factors Affecting Monthly Earnings

Let us examine a data set that explores the relationship between total monthly earnings (`MonthlyEarnings`) and a number of factors that may influence monthly earnings including including each person's IQ (`IQ`), a measure of knowledge of their job (`Knowledge`), years of education (`YearsEdu`), years experience (`YearsExperience`), and years at current job (`Tenure`).

The code below downloads a CSV file that includes data on the above variables from 1980 for 935 individuals and assigns it to a data set that we name `wages`.

```
wages <- read.csv("http://murraylax.org/datasets/wage2.csv");
```

The following call to `lm()` estimates a multiple regression predicting monthly earnings based on the five explanatory variables given above. The next call to `summary()` displays some summary statistics for the estimated regression.

```
lmwages <- lm(MonthlyEarnings
              ~ IQ + Knowledge + YearsEdu + YearsExperience + Tenure,
              data = wages)
summary(lmwages)
```

```
##
## Call:
## lm(formula = MonthlyEarnings ~ IQ + Knowledge + YearsEdu + YearsExperience +
##     Tenure, data = wages)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -826.33 -243.85  -44.83  180.83 2253.35
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -531.0392   115.0513  -4.616 4.47e-06 ***
## IQ                 3.6966     0.9651   3.830 0.000137 ***
## Knowledge          8.2703     1.8273   4.526 6.79e-06 ***
## YearsEdu          47.2698     7.2980   6.477 1.51e-10 ***
## YearsExperience   11.8589     3.2494   3.650 0.000277 ***
## Tenure             6.2465     2.4565   2.543 0.011156 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 365.4 on 929 degrees of freedom
## Multiple R-squared:  0.1878, Adjusted R-squared:  0.1834
## F-statistic: 42.97 on 5 and 929 DF,  p-value: < 2.2e-16
```

Notice that the regression predict the return to each new year of experience on monthly earnings is equal to \$11.86. The return to an additional year of education is equal to \$47.27.

Is it possible that the return to additional years of experience should be bigger for people with higher levels of education? Could this be another benefit of receiving a higher education that is not yet in our model? It could be that not only do people with a higher level of education benefit with a higher base pay, the rates of pay *increases* for additional experience with these types of jobs may be higher.

We introduce an interaction effect between experience and education to allow for this effect and estimate its significance. We introduce a new term into the regression that is equal to the `YearsEdu` variable *multiplied by* the `YearsExperience` variable. The following call to `lm()` estimates such a model:

```
lmwages <- lm(MonthlyEarnings
              ~ IQ + Knowledge + YearsEdu + YearsExperience + Tenure
              + YearsEdu*YearsExperience,
              data = wages)
summary(lmwages)
```

```
##
## Call:
## lm(formula = MonthlyEarnings ~ IQ + Knowledge + YearsEdu + YearsExperience +
##     Tenure + YearsEdu * YearsExperience, data = wages)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -837.38 -239.36  -42.32  182.21 2239.47
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -62.200    232.918  -0.267 0.789491
## IQ                           3.589      0.964   3.723 0.000209 ***
## Knowledge                    7.996      1.827   4.377 1.34e-05 ***
## YearsEdu                    13.213     16.424   0.804 0.421336
## YearsExperience            -30.771     18.711  -1.645 0.100406
## Tenure                       6.649      2.457   2.706 0.006929 **
## YearsEdu:YearsExperience     3.307      1.430   2.313 0.020924 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 364.5 on 928 degrees of freedom
## Multiple R-squared:  0.1925, Adjusted R-squared:  0.1873
## F-statistic: 36.87 on 6 and 928 DF,  p-value: < 2.2e-16
```

The coefficient on the interaction term is equal to 3.307 and it is statistically significant at the 5% level (p-value = 0.0209). We found sufficient statistical evidence that experience and education have an interaction effect.

The coefficient is positive which implies higher levels of education lead to higher return for an additional year of experience. It is equivalent to say that higher levels of experience lead to a higher return to an additional year of education.

2. Marginal Effects

What is the impact on monthly earnings of an additional year of experience? This is called the *marginal effect* of a year of experience

In the simple linear regression with no interaction effects at the opening of this tutorial, the marginal effect is exactly the same as the estimate of the coefficient on experience

With interaction effects, you can no longer answer the question looking at only one coefficient. When experience increases by one year, the coefficient on `YearsExperience` suggests that monthly earnings *decreases* by \$30.77. *That is not the end of the story.*

The coefficient on the interaction effect says that an additional year of experience leads to a \$3.31 raise *for each year of education* the person has obtained. *The marginal effect is different for each person, depending on that person's average level of education.*

The marginal effect for experience is given by,

$$me = b_{Exp} + b_{Exp*Edu} x_{Edu}$$

For someone with 16 years of education (most likely a four-year college degree), the marginal effect for experience is equal to:

$$me = -30.77 + 3.31(16) = \$22.19$$

The marginal effect on monthly earnings of an additional year of experience for a four year college grad is \$22.19.

Let us examine the marginal effect for someone with 12 years of education (most likely someone with a high school degree and no college):

$$me = -30.77 + 3.31(12) = \$8.95$$

To compute the *average marginal effect*, we can replace the mean value for years of education into the formula above:

$$\bar{me} = b_{Exp} + b_{Exp*Edu} \bar{x}_{Edu}$$

Let us compute this in R:

```
b <- lmwages$coefficients;
me_exp <- b["YearsExperience"] + b["YearsEdu:YearsExperience"] * mean(wages$YearsEdu)
me_exp
```

```
## YearsExperience
##        13.77451
```

The average return to an additional year of experience is equal to \$13.77