Profesor: Héctor Bahamonde, PhD.

e:hector.bahamonde@uoh.cl

w:www.hectorbahamonde.com

Curso: MLE.

TA: Gonzalo Barría.

Términos de Interacción: Introducción

Un modelo lineal tradicional—como el de la Equation 1—te muestra el efecto de una variable x_1 (*educación*) sobre y manteniendo la variable de control x_2 (*hombre*) constante en su media.

$$\text{ingresos}_i \;=\; \beta_0 + \beta_1 \text{educación}_i + \beta_2 \text{hombre}_i + \epsilon_i \tag{1}$$

Por qué es importante controlar por género?

Un término de interacción, sin embargo, se usa cuando queremos saber el efecto **combinado** de dos variables. La ventaja del término de interacción es que muestra el efecto de las dos variables x_1 **y** x_2 impactando y al mismo tiempo. Por ejemplo, si quisiéramos saber cuál es el efecto que tiene la variable *educación* (x_1) **y** (esto es, en **combinación con**) *hombre* (x_2) sobre *ingresos* (y_i), deberíamos estimar la siguiente ecuación:

$$\text{ingresos}_i \;=\; \beta_0 + \beta_1 \text{educación}_i + \beta_2 \text{hombre}_i + \beta_3 \text{educación}_i \times \text{hombre}_i + \epsilon_i \tag{2}$$

En este ejemplo asume que la variable *hombre* sólo toma valores 0 (mujer) y 1 (hombre).

Substantivo Desde un punto de vista substantivo los terminós de interacción son relevantes porque dan luz a preguntas de tipo interactivas: existen serias razones para pensar que—lamentablemente—no es lo mismo ser un hombre con educación a una mujer con educación. Nota cómo ahora estamos refiriéndonos a *ambos* factores impactando la variable dependiente de manera conjunta.

Parametrización La manera en la que está parametrizada el modelo lineal con interacciones es *parecida* al modelo tradicional. Pero existen algunas diferencias. Volvamos a la Equation 2.

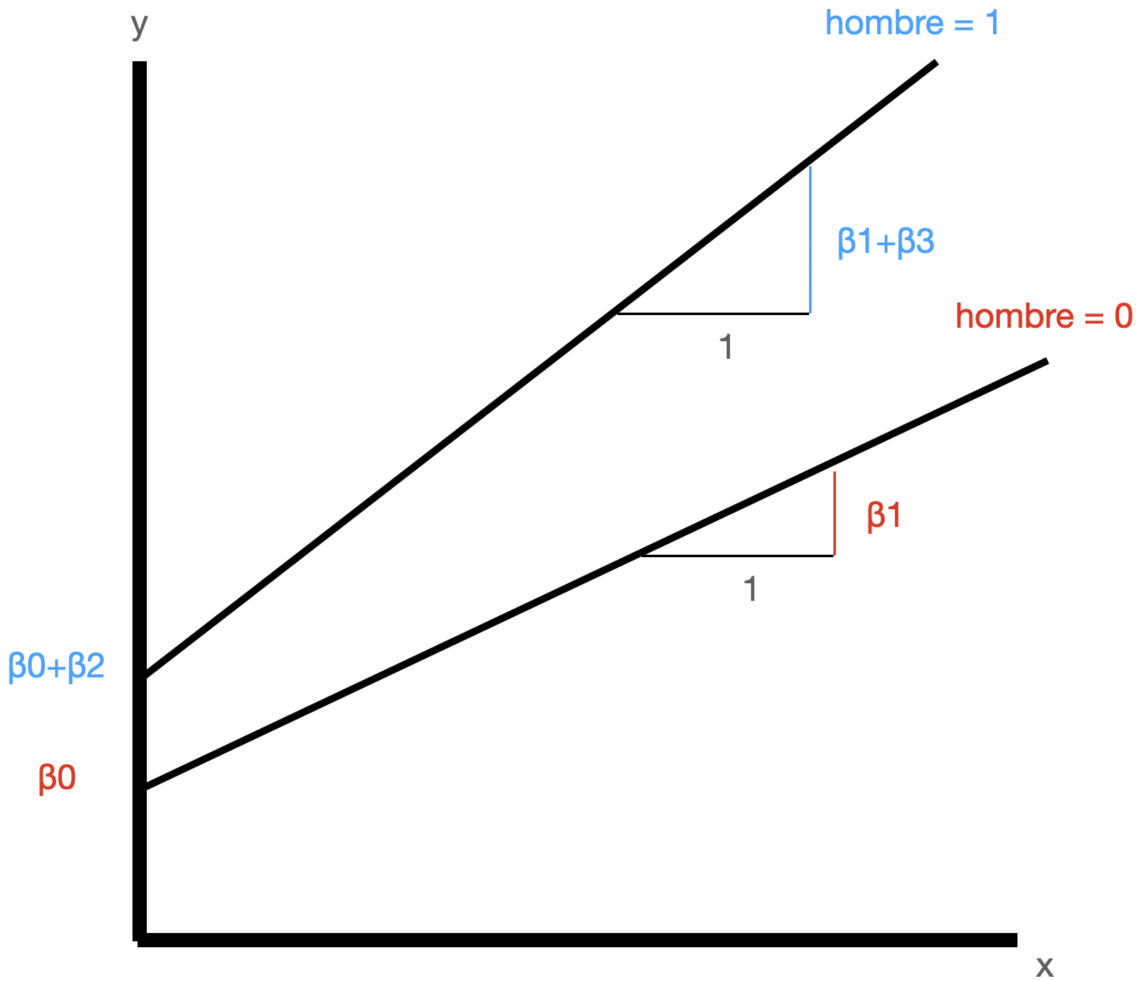
1. Los parámetros β_0 y β_1 son el intercepto y la pendiente—respectivamente—de la "categoría de referencia", es decir, de la categoría base: cuando la variable *hombre*=0 (es decir, el caso

de las mujeres). Esto es muy intuitivo. Cuando la variable *hombre*=0, el modelo se reduce a Equation 3.

2. El intercepto para el otro grupo—*hombre*=1, es decir, el caso de los hombres—es $\beta_0 + \beta_2$, mientras que la pendiente está dada por $\beta_1 + \beta_3$. De la misma manera, es muy intuitivo. Ve la Equation 4

$$
\begin{aligned}
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + \beta_2 \text{hombre}_i + \beta_3 \text{educación}_i \times \text{hombre}_i + \epsilon_i \\
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + \beta_2 0 + \beta_3 \text{educación}_i \times 0_i + \epsilon_i \\
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + 0 + 0 + \epsilon_i \\
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + \epsilon_i
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + \beta_2 \text{hombre}_i + \beta_3 \text{educación}_i \times \text{hombre}_i + \epsilon_i \\
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + \beta_2 1 + \beta_3 \text{educación}_i \times 1_i + \epsilon_i \\
\text{ingresos}_i &= \beta_0 + \beta_1 \text{educación}_i + \beta_2 + \beta_3 \text{educación}_i + \epsilon_i \\
\text{ingresos}_i &= (\beta_0 + \beta_2) + \text{educación}_i \times (\beta_1 + \beta_3) + \epsilon_i
\end{aligned}
\tag{4}
$$



Buenas Prácticas

Fíjate que cada vez que incluimos un término de interacción ($\text{educación}_i \times \text{hombre}_i$), para interpretar su parámetro asociado (β_3), es necesario incluir los sub-términos por separado. Esto es, permitir que la ecuación tenga un parámetro independiente asociado a *educación* y *hombre*, esto es, β_1 y β_2 (tal y como aparece en Equation 2). Si estimamos sólo la siguiente ecuación, β_3 estará sesgado. **NO HAGAS LO SIGUIENTE**:

$$\text{ingresos}_i = \beta_0 + \beta_3 \text{hombre}_i \times \text{educación}_i + \epsilon_i \tag{5}$$

Estimación en R

De acuerdo a Brambor, Clark, and Golder (2006, p. 73),[1] el efecto marginal en la ecuación Equation 6,

$$\text{ingresos}_i \;=\; \beta_0 + \beta_1 \text{educación}_i + \beta_2 \text{hombre}_i + \epsilon_i \tag{6}$$

está dado por el siguiente cálculo:

$$\frac{\partial y}{\partial x_1} = \beta_1 \tag{7}$$

Sin embargo el término de interacción en Equation 2 es distinto, y está dado por el siguiente cálculo:

$$\frac{\partial y}{\partial x_1} = \beta_1 + \beta_3 \text{hombre} \tag{8}$$

En palabras, es *cuánto cambia y cuando cambia x, según niveles de la variable hombre.*

Cambiemos de ejemplo, y estimemos un modelo con tres niveles (no dos, como *hombre*).

Carguemos los datos:

```
p_load(effects)
data(Duncan)
summary(Duncan)

##    type        income        education        prestige
##  bc  :21   Min.   : 7.00   Min.   :  7.00   Min.   : 3.00
##  prof:18   1st Qu.:21.00   1st Qu.: 26.00   1st Qu.:16.00
##  wc  : 6   Median :42.00   Median : 45.00   Median :41.00
##            Mean   :41.87   Mean   : 52.56   Mean   :47.69
##            3rd Qu.:64.00   3rd Qu.: 84.00   3rd Qu.:81.00
##            Max.   :81.00   Max.   :100.00   Max.   :97.00
```

Estimemos el modelo. Nota que hemos puesto la multiplicación, y `R` "sabe" que debe meter los términos constitutivos.

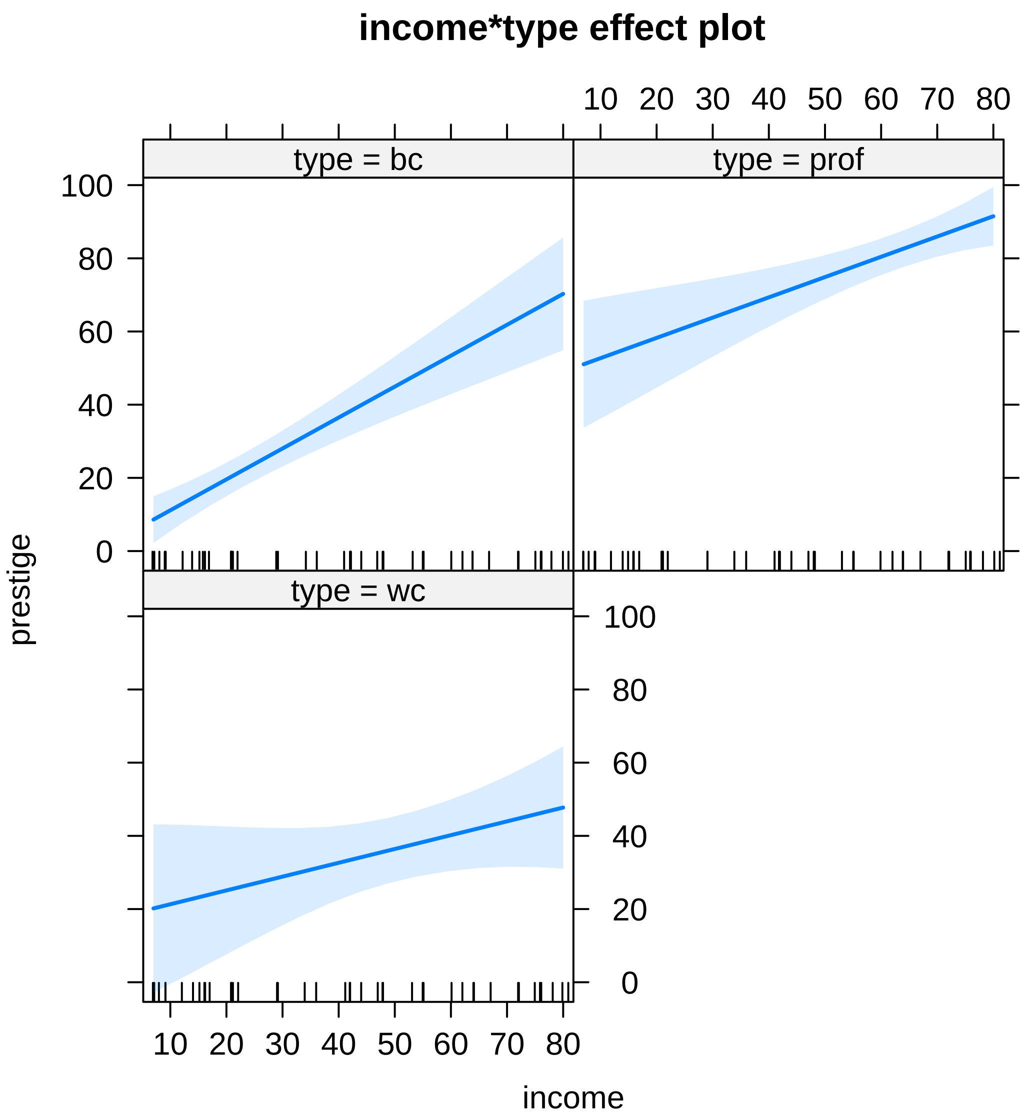
[1] Ecuaciones 11-13.

```
modelo.1 = lm(prestige ~ income*type, data = Duncan)
summary(modelo.1)
```

```
##
## Call:
## lm(formula = prestige ~ income * type, data = Duncan)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.4405  -6.0480  -0.2787   4.7269  28.1950
##
## Coefficients:
##                  Estimate Std. Error t value     Pr(>|t|)
## (Intercept)        2.6828     3.8182   0.703     0.486450
## income             0.8450     0.1289   6.554 0.0000000882 ***
## typeprof          44.4868    10.3641   4.292     0.000113 ***
## typewc            14.8531    13.4986   1.100     0.277926
## income:typeprof   -0.2909     0.2017  -1.442     0.157148
## income:typewc     -0.4674     0.2736  -1.709     0.095466 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.44 on 39 degrees of freedom
## Multiple R-squared:  0.9026,Adjusted R-squared:  0.8902
## F-statistic: 72.31 on 5 and 39 DF,  p-value: < 0.00000000000000022
```

Como explican Brambor, Clark, and Golder (2006), las tablas de regresión no nos ayudan a interpretar los modelos interactivos. Debemos proceder interpretando como se señala en Equation 8. Afortunadamente existe la librería `effects`.

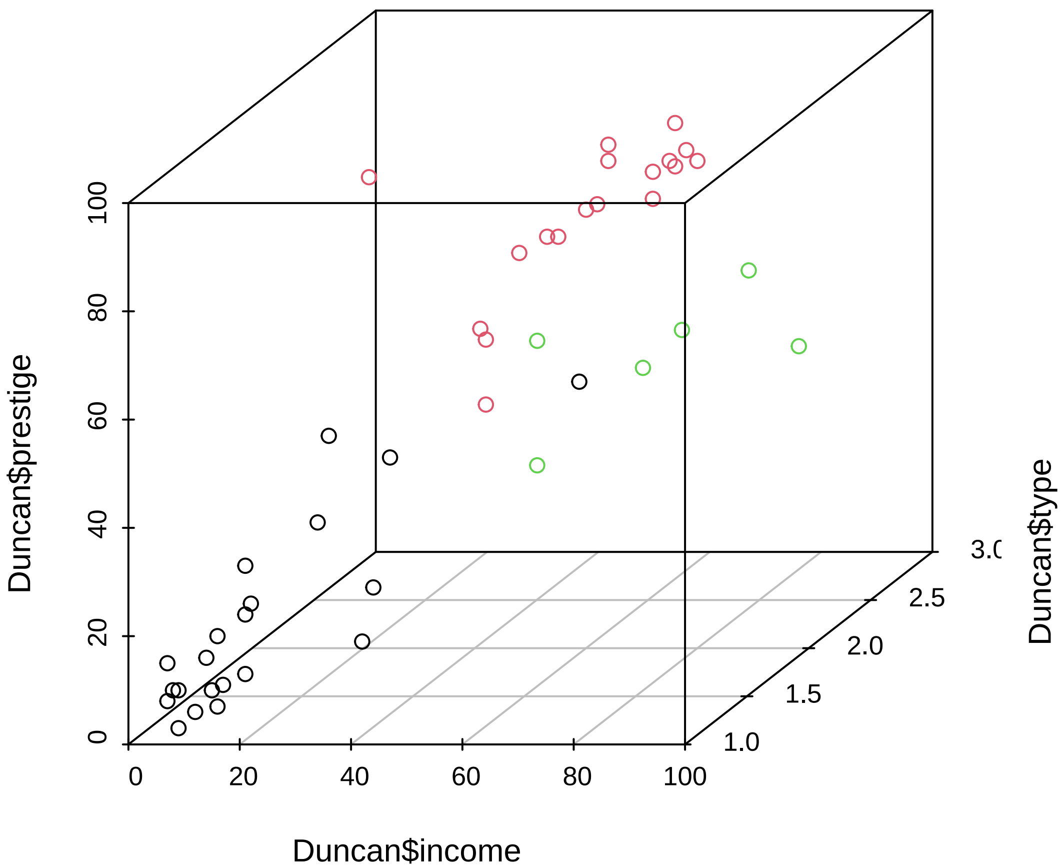
```
term.int <- effect("income*type", modelo.1)
plot(term.int, as.table=T)
```



Si te das cuenta, los efectos no son los mismos. Las derivadas (en Equation 8) no tienen por qué dar lo mismo. Es por esto que no debemos mirar la tabla de regresión. En un sentido espacial, un término de interacción es el análisis de tres planos. En la Equation 2: y, x_1, x_2.

Veamos de qué se trata:

```
p_load(scatterplot3d)
scatterplot3d(Duncan$income, Duncan$type, Duncan$prestige, color = as.numeric(Duncan$type))
```



Usemos una base de datos donde todas las variables son continuas (no como en el ejemplo donde *hombre* es dicotómica):

```
p_load(car,rgl)
data(iris)
sep.l <- iris$Sepal.Length
sep.w <- iris$Sepal.Width
pet.l <- iris$Petal.Length
scatter3d(x = sep.l, y = pet.l, z = sep.w, groups = iris$Species)
```

Correr esto último en `R`.

```
## [1] "Clase_14.R"
## Writing to file Clase_14.R
```