

Interacción entre variables

Regresión lineal con interacciones entre variable
dicotómicas

PSI4035

Metodología Cuantitativa Avanzada I

Abril 27
2022

Carrasco, D., PhD,
Centro de Medición MIDE UC
Pontificia Universidad Católica de Chile

Metodología Cuantitativa

Elementos de una regresión

Repaso de componentes de una regresión

Regresión lineal entre variables continuas

22

Chapter 1 Review of linear regression

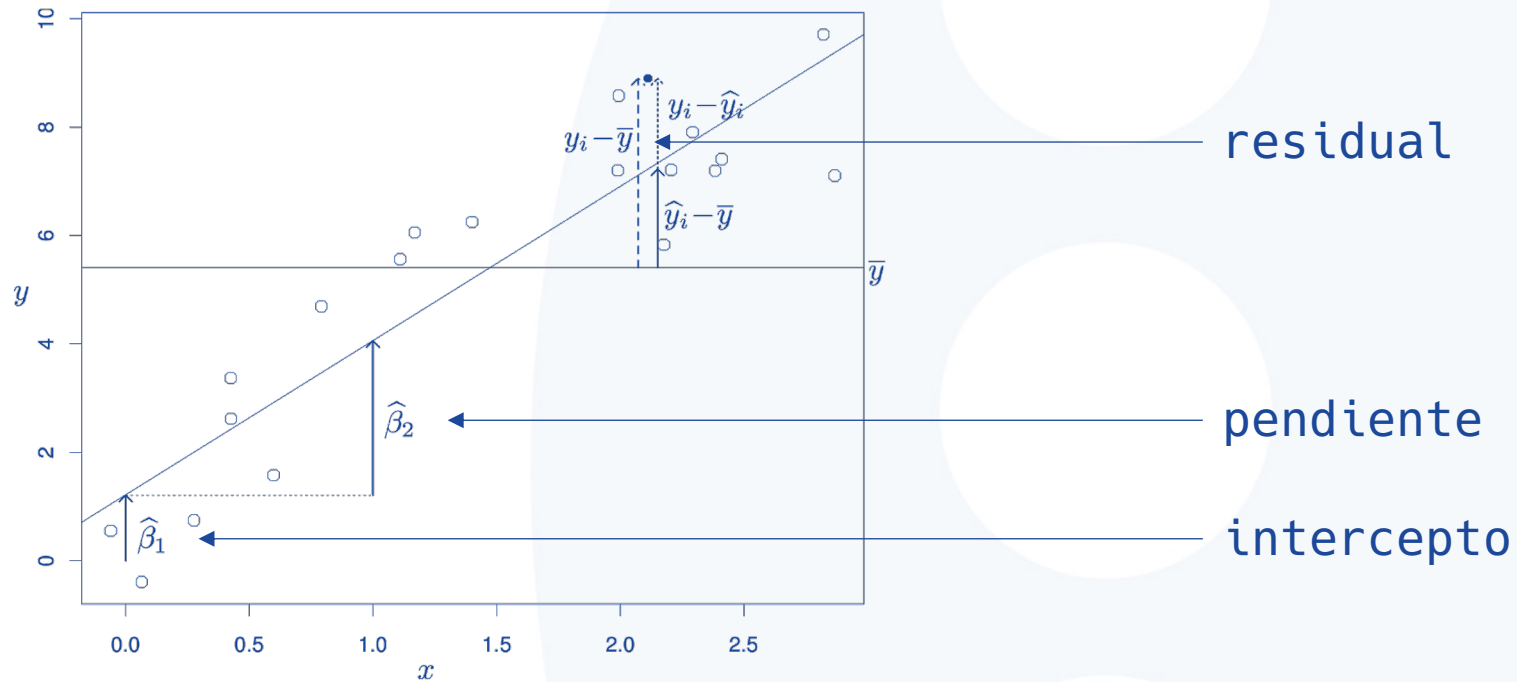


Figure 1.5: Illustration of sums of squares for simple linear regression

Regresión lineal empleando una covariable dicotómica

28

Chapter 1 Review of linear regression

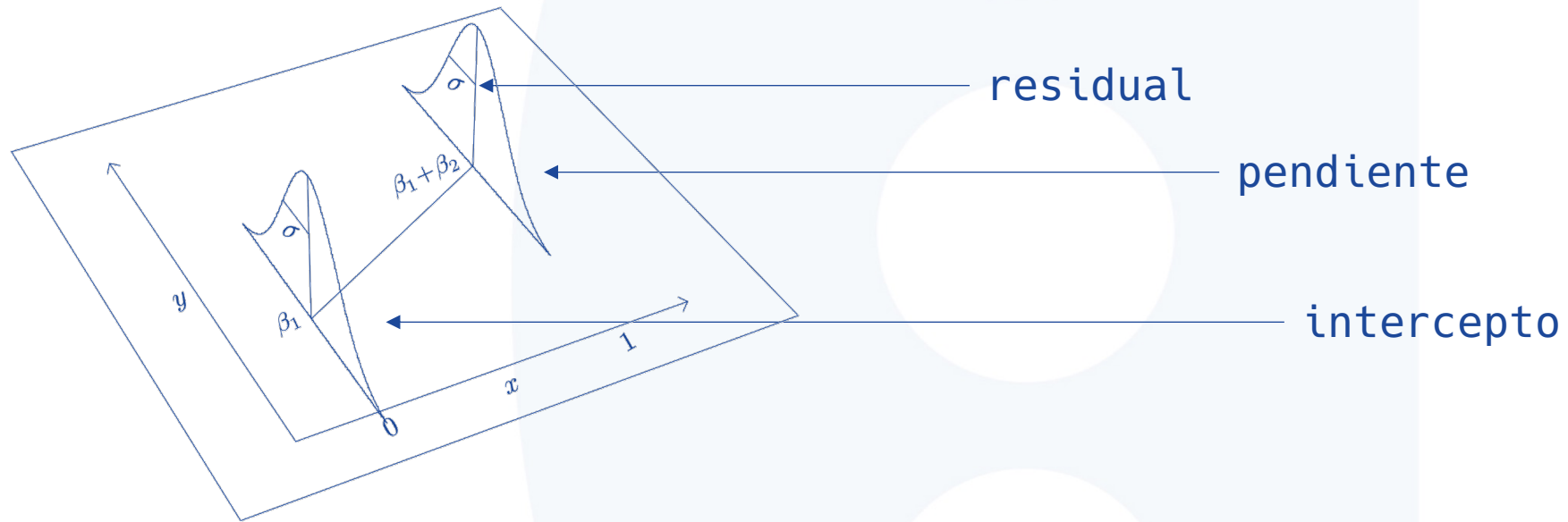


Figure 1.7: Illustration of simple linear regression with a dummy variable

Modelos de regresión

Existen diferentes modelos de regresión además de la regresión lineal:

Regresión lineal

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Regresión logística

$$\text{logit}\{\Pr(y_i = 1|x_i)\} = \alpha + \beta x_i$$

Regresión ordinal (modelo generalizado)

$$y_i^* = \alpha + \beta x_i + \epsilon_i$$

Modelo Multinivel

$$y_{ij} = \alpha + \beta x_{ij} + \mu_j + \epsilon_{ij}$$

Regresión latente

$$\eta_i = \alpha + \beta x_i + \epsilon_i$$

Ejemplos de código de cada tipo de regresión:

Regresión lineal

```
lm(y ~ 1 + x, data = data_example)
```

Regresión logística

```
glm(y ~ 1 + x, data = data_example, family = binomial)
```

Regresión ordinal (modelo generalizado)

```
ordinal::clm(y ~ 1 + x, data = data_example)
```

Modelo Multinivel

```
lme4::lmer(y ~ (1|id_j) + x, data = data_example)
```

Regresión latente

```
# more complicated syntax
```

Metodología Cuantitativa

Regresión

Interacciones entre variables dicotómicas

Interacciones con una variable dicotómica

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. 21: 489–519 (2006)
Published online 8 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.851

ESTIMATING THE EFFECT OF SMOKING ON BIRTH OUTCOMES USING A MATCHED PANEL DATA APPROACH

JASON ABREVAYA*

Department of Economics, Purdue University, West Lafayette, IN 47907-2056, USA

SUMMARY

Estimating the casual effect of smoking on birth outcomes is difficult since omitted (unobserved) variables are likely to be correlated with a mother's decision to smoke. While some previous work has dealt with this endogeneity problem by using instrumental variables, this paper instead attempts to estimate the smoking effect from panel data (i.e., data on mothers with multiple births). Panel data sets are constructed with matching algorithms applied to federal natality data. The fixed effects regressions, which control for individual heterogeneity, yield significantly different results from ordinary least squares and previous instrumental variable approaches. The potential inconsistency caused by 'false matches' and other violations of the fixed effects strict exogeneity assumption are considered. Copyright © 2006 John Wiley & Sons, Ltd.

1. INTRODUCTION

A 2001 report by the Surgeon General summarizes the relationship between smoking during pregnancy and birth outcomes as follows: 'Infants born to women who smoke during pregnancy have a lower average birthweight and are more likely to be small for gestational age than infants born to women who do not smoke. Low birthweight is associated with increased risk for neonatal, perinatal, and infant morbidity and mortality. The longer the mother smokes during pregnancy, the greater the effect on the infant's birthweight. . . [S]moking is also associated with a modest increase in risk for preterm delivery.' (*Women and Smoking: A Report of the Surgeon General*, Centers for Disease Control and Prevention, 2001).

The issue of low birthweight has received a great deal of attention in the economics literature due to the direct medical costs and long-term costs associated with low birthweight. While low-birthweight (LBW) infants (defined as weighing less than 2500 g at birth) account for less than 10% of births in the United States, Lewit *et al.* (1995) estimate that LBW infants account for more than one-third of health care costs during the first year of life. Joyce (1999) estimates that the marginal cost of hospital newborn care is \$6–10/g, with higher marginal costs at lower birthweights. Several studies have shown that lower birthweight is associated with higher infant mortality rates (see, for instance, the references in Mathews (2001)). Schoendorf and Kiely (1992) link maternal smoking to higher rates of sudden infant death syndrome.

Recent research has also shown that, in addition to the short-term costs, low birthweight can have substantial long-term effects as well. Hack *et al.* (1995) find that low-birthweight babies have developmental problems in cognition, attention and neuromotor functioning that persist

* Correspondence to: Jason Abrevaya, Department of Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907-2056, USA. E-mail: abrevaya@purdue.edu
Contract/grant sponsor: John & Mary Willis Young Faculty Scholar Fund, Purdue University.
Contract/grant sponsor: Faculty research Fund, University of Chicago GSB.

- Para ilustrar una interacción, vamos a emplear datos de Abrevaya (2006). Estos son una serie de registros del peso al nacer de los hijos de diferentes madres. La motivación del estudio, es evaluar el efecto que posee el fumar en el peso de los nacidos. Este ejemplo se encuentra discutido con mayor profundidad en Rabe-Hesketh & Skrondal (2012), y consiste en un 10% de los datos originales ($n = 3978$ madres).
- A diferencia de los ejemplos de las clases previas, vamos a emplear más predictores. Es decir que no solo veremos como interpretar una regresión simple, con un solo predictor, sino que con varios predictores.
- La interpretación no cambia tanto. En el caso de la regresión múltiple, nos sigue interesando los valores esperados de y_i , condicionales a los valores de x_i y otras covariables.
- Primero vamos a evaluar la interacción de dos variables dicotómicas.

Interacciones con una variable dicotómica

```
#-----  
# prepare data  
#-----  
  
library(dplyr)  
data_first <- psi2301::smoking %>%  
  dplyr::filter(idx == 1) %>%  
  mutate(kg = birwt/1000) %>%  
  mutate(mage_g = mean(mage, na.rm = TRUE)) %>%  
  mutate(mage_m = mage - mage_g) %>%  
  r4sda::remove_labels()  
  
#-----  
# variables  
#-----  
  
# kg          # peso al nacer en kilogramos  
# novisit     # sin visita de control prenatal (1 = sin visitas, 0 = con visitas)  
# smoke       # fuma  
  
# -----  
# regression in steps  
# -----  
  
m00 <- as.formula('kg ~ 1')  
m01 <- as.formula('kg ~ 1 + smoke')  
m02 <- as.formula('kg ~ 1 + novisit')  
m03 <- as.formula('kg ~ 1 + novisit*smoke')  
  
f00 <- lm(m00, data = data_first)  
f01 <- lm(m01, data = data_first)  
f02 <- lm(m02, data = data_first)  
f03 <- lm(m03, data = data_first)  
  
texreg::screenreg(  
  list(f00, f01, f02, f03),  
  star.symbol = "*",  
  center = TRUE,  
  doctype = FALSE,  
  dcolumn = TRUE,  
  booktabs = TRUE,  
  single.row = FALSE  
)
```

- Para llamar a los datos empleamos el comando `psi231::smoking`
- Seleccionamos los datos del primer hijo, de modo que tengamos datos independientes. En caso contrario, tendríamos datos de más de un nacido de una misma madre, y estaríamos violando el supuesto de independencia de las observaciones.
- Vamos a emplear solo tres medidas: el peso al nacer, si la madre fuma (0 = no fuma; 1 = si fuma), y si ha faltado a los controles pre natales (1 = sin visita, 0 = con visitas).
- El peso de los recién nacidos está en gramos, así que lo convertiremos a kilos, de modo que sea más fácil de interpretar (i.e. leer los estimados en Kg.).
- Luego ajustamos los modelos de regresión en pasos. Primero solo la covariable de interés (smoke). Luego, el posible moderador (novisit), y finalmente un modelo con ambos predictores y su interacción.

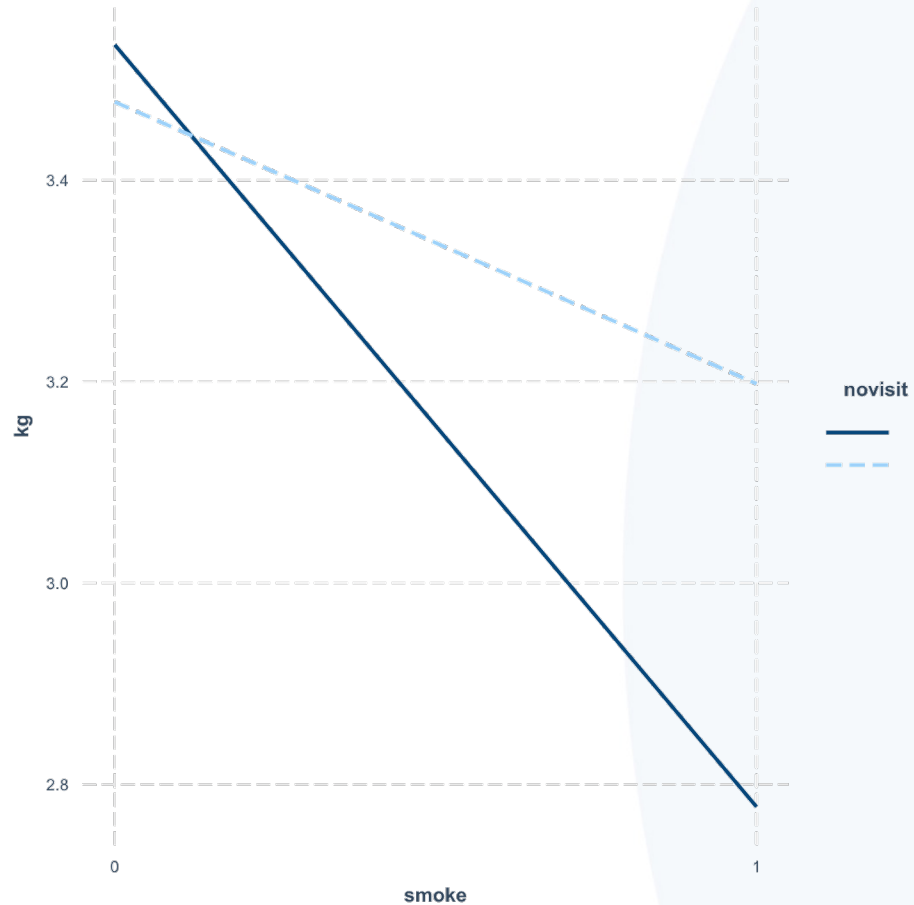
Interacciones con una variable dicotómica

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3.44 *** (0.01)	3.48 *** (0.01)	3.44 *** (0.01)	3.48 *** (0.01)
smoke		-0.29 *** (0.02)		-0.28 *** (0.02)
novisit			-0.23 * (0.09)	0.06 (0.12)
novisit:smoke				-0.48 ** (0.18)
R ²	0.00	0.04	0.00	0.04
Adj. R ²	0.00	0.04	0.00	0.04
Num. obs.	3978	3978	3978	3978

*** p < 0.001; ** p < 0.01; * p < 0.05

- Nuestra variable dependiente es el peso del primer nacido de cada mujer que participa en el estudio, en kilogramos.
- En promedio, los recién nacidos presentan 3.44 kg al nacer (ver Model 1).
- Entre los nacidos de madres fumadoras, se observa una diferencia de .29 kg, en contraste a los nacidos de madres no fumadoras ($b = -0.29$, $SE = .02$, $p < .001$), factor que explica hasta un 4% de la varianza del peso de los nacidos.
- Adicionalmente, en el modelo 3, se observa que los nacidos de madres que no asisten a control pre-natal (novisit), también presentan una diferencia negativa en peso al nacer de .23 kg, en contraste a los nacidos de madres que sí asisten a los controles prenatales ($b = -0.23$, $SE = .09$, $p < .05$),
- Finalmente, en el modelo 4, observamos un efecto de interacción. Los hijos de madres fumadoras, que no asisten a los controles natales presentan mayores diferencias en peso, en contraste al resto hijo de madres fumadoras ($b = -0.48$, $SE = .18$, $p < .01$). Este ultimo modelo, explica hasta un 4% de la varianza del peso de los recién nacidos ($R^2 = .04$, $F(3,3874) = 57.90$, $p < .001$).

Interacciones con una variable dicotómica



- Nuestra variable dependiente es el peso del primer nacido de cada mujer que participa en el estudio, en kilogramos.
- En promedio, los recién nacidos presentan 3.44 kg al nacer (ver Model 1).
- Entre los nacidos de madres fumadoras, se observa una diferencia de .29 kg, en contraste a los nacidos de madres no fumadoras ($b = -0.29$, $SE = .02$, $p < .001$), factor que explica hasta un 4% de la varianza del peso de los nacidos.
- Adicionalmente, en el modelo 3, se observa que los nacidos de madres que no asisten a control pre-natal (novisit), también presentan una diferencia negativa en peso al nacer de .23 kg, en contraste a los nacidos de madres que sí asisten a los controles prenatales ($b = -0.23$, $SE = .09$, $p < .05$),
- Finalmente, en el modelo 4, observamos un efecto de interacción. Los hijos de madres fumadoras, que no asisten a los controles natales presentan mayores diferencias en peso, en contraste al resto hijo de madres fumadoras ($b = -0.48$, $SE = .18$, $p < .01$). Este ultimo modelo, explica hasta un 4% de la varianza del peso de los recién nacidos ($R^2 = .04$, $F(3,3874) = 57.90$, $p < .001$).

Interacciones con una variable dicotómica

Residuals:

Min	1Q	Median	3Q	Max
-2.52762	-0.30262	0.00938	0.32138	1.79213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.477620	0.008508	408.741	<2e-16 ***
novisit	0.056733	0.120623	0.470	0.6381
smoke	-0.279750	0.022837	-12.250	<2e-16 ***
novisit:smoke	-0.476680	0.184205	-2.588	0.0097 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4961 on 3974 degrees of freedom
Multiple R-squared: 0.04188, Adjusted R-squared: 0.04116
F-statistic: 57.9 on 3 and 3974 DF, p-value: < 2.2e-16

- Nuestra variable dependiente es el peso del primer nacido de cada mujer que participa en el estudio, en kilogramos.
- En promedio, los recién nacidos presentan 3.44 kg al nacer (ver Model 1).
- Entre los nacidos de madres fumadoras, se observa una diferencia de .29 kg, en contraste a los nacidos de madres no fumadoras ($b = -0.29$, $SE = .02$, $p < .001$), factor que explica hasta un 4% de la varianza del peso de los nacidos.
- Adicionalmente, en el modelo 3, se observa que los nacidos de madres que no asisten a control pre-natal (novisit), también presentan una diferencia negativa en peso al nacer de .23 kg, en contraste a los nacidos de madres que sí asisten a los controles prenatales ($b = -0.23$, $SE = .09$, $p < .05$),
- Finalmente, en el modelo 4, observamos un efecto de interacción. Los hijos de madres fumadoras, que no asisten a los controles natales presentan mayores diferencias en peso, en contraste al resto hijo de madres fumadoras ($b = -0.48$, $SE = .18$, $p < .01$). Este último modelo, explica hasta un 4% de la varianza del peso de los recién nacidos ($R^2 = .04$, $F(3,3874) = 57.90$, $p < .001$).

Metodología Cuantitativa

Regresión con interacción

Hipótesis e interpretación

Interacciones con una variable dicotómica: formulación de hipótesis.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3.44 *** (0.01)	3.48 *** (0.01)	3.44 *** (0.01)	3.48 *** (0.01)
smoke		-0.29 *** (0.02)		-0.28 *** (0.02)
novisit			-0.23 * (0.09)	0.06 (0.12)
novisit:smoke				-0.48 ** (0.18)
R^2	0.00	0.04	0.00	0.04
Adj. R^2	0.00	0.04	0.00	0.04
Num. obs.	3978	3978	3978	3978

*** p < 0.001; ** p < 0.01; * p < 0.05

- En una hipótesis de interacción, tenemos la sospecha de que un moderador influye sobre la relación entre x e y. Es decir, que los valores esperados de y, debido a x, no son constantes, sino que pueden verse aumentado o disminuidos debido a la variable moderadora (llamémosla w).
- En este caso, estamos diciendo que podemos explicar a la variable respuesta "y", condicional a los valores de "x" y "w", y que además la relación de "yx" esta influida por los valores "w".
- Esto quiere decir que, si nuestro modelo es:
- $y_i = \alpha + \beta x_i + \gamma w_i + \delta x_i w_i + \epsilon_i$
- Nuestra hipótesis nula es que $\delta = 0$, mientras que nuestra hipótesis alternativa es que $\delta \neq 0$. En palabras, que la diferencia en el peso esperado de los recién nacidos, entre las madres que fuman y no fuman, son similares tanto para las madres que asisten a los controles prenatales, como para las madres que no asisten a los controles prenatales.
- Nuestra hipótesis alternativa, es que la diferencia de peso esperado de los recién nacidos, entre las madres que fuman y no fuman, son diferentes según las madres asistan o no, a los controles prenatales.

Interacciones con una variable dicotómica: código

```
#-----  
# prepare data  
#-----  
  
library(dplyr)  
data_first <- psi2301::smoking %>%  
  dplyr::filter(idx == 1) %>%  
  mutate(kg = birwt/1000) %>%  
  mutate(mage_g = mean(mage, na.rm = TRUE)) %>%  
  mutate(mage_m = mage - mage_g) %>%  
  r4sda::remove_labels()  
  
#-----  
# variables  
#-----  
  
# kg          # peso al nacer en kilogramos  
# novisit     # sin visita de control prenatal (1 = sin visitas, 0 = con visitas)  
# smoke       # fuma  
  
# -----  
# codigo 1  
# -----  
  
m03 <- as.formula('kg ~ 1 + novisit*smoke')  
f03 <- lm(m03, data = data_first)  
summary(f03)  
  
# -----  
# codigo 2  
# -----  
  
m03 <- as.formula('kg ~ 1 + novisit + smoke + novisit*smoke')  
f03 <- lm(m03, data = data_first)  
summary(f03)
```

- R permite ajustar modelos de interacción ya sea creando los componentes de interacción sobre la base de datos, o empleando el símbolo "*" entre dos covariables.

- Esto quiere decir que el usuario puede escribir:

$$\text{lm}(y \sim x*w)$$

- Con lo cual "lm" asume:

$$y_i = \alpha + \beta x_i + \gamma w_i + \delta x_i w_i + \epsilon_i$$

- Otra alternativa es escribir el código completo:

$$\text{lm}(y \sim x + w + x*w)$$

- Ambos códigos debieran entregar los mismos resultados.

Interacciones con una variable dicotómica: significancia del modelo

Residuals:

Min	1Q	Median	3Q	Max
-2.52762	-0.30262	0.00938	0.32138	1.79213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.477620	0.008508	408.741	<2e-16 ***
novisit	0.056733	0.120623	0.470	0.6381
smoke	-0.279750	0.022837	-12.250	<2e-16 ***
novisit:smoke	-0.476680	0.184205	-2.588	0.0097 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4961 on 3974 degrees of freedom

Multiple R-squared: 0.04188, Adjusted R-squared: 0.04116

F-statistic: 57.9 on 3 and 3974 DF, p-value: < 2.2e-16

- Hay dos formas generales de evaluar globalmente un modelo de regresión: la prueba F, y los indicadores de R^2 .
- El primero nos indica que los factores ingresados al modelo, explican o condicionan las medias esperadas de la variable de respuesta. Al igual que en el caso de ANOVA, nos indican si los factores incluidos en el modelo producen un componente de varianza que se encuentre dentro de lo esperado, o si estos factores producen un valor F atípico. Cuando es el segundo caso, nuestro valor F será mayor al punto crítico, y nuestro p valor debiera ser menor a $p < .05$; $p < .01$ ó $p < .001$. Al igual que en el caso convencional de ANOVA, se incluyen los grados de libertad, se reporta el F obtenido, y el p valor observado: $F(3,3874) = 57.90, p < .001$
- El segundo indicador nos indica la porción de varianza explicada por los factores incluidos en el modelo. En este caso, es 4% ($R^2 = .04$). Una manera de obtener este valor, es generar los valores predichos, y correlacionarlos con los datos observados.

Interacciones con una variable dicotómica: intercepto

Residuals:

Min	1Q	Median	3Q	Max
-2.52762	-0.30262	0.00938	0.32138	1.79213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.477620	0.008508	408.741	<2e-16 ***
novisit	0.056733	0.120623	0.470	0.6381
smoke	-0.279750	0.022837	-12.250	<2e-16 ***
novisit:smoke	-0.476680	0.184205	-2.588	0.0097 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4961 on 3974 degrees of freedom

Multiple R-squared: 0.04188, Adjusted R-squared: 0.04116

F-statistic: 57.9 on 3 and 3974 DF, p-value: < 2.2e-16

- ¿Qué significa el intercepto cuando hay más variables? Es la media esperada de la variable de respuesta, cuando todas las covariables incluidas en el modelo son cero.
- En este caso, es el peso esperado de un recién nacido, cuando la madre “no fuma” (smoke == 0), y cuando “asiste a los controles natales” (novisit == 0). En general, esperamos que un recién nacido con madres de estas características pese 3.48 kg al nacer.

Interacciones con una variable dicotómica: pendiente

Residuals:

Min	1Q	Median	3Q	Max
-2.52762	-0.30262	0.00938	0.32138	1.79213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.477620	0.008508	408.741	<2e-16 ***
novisit	0.056733	0.120623	0.470	0.6381
smoke	-0.279750	0.022837	-12.250	<2e-16 ***
novisit:smoke	-0.476680	0.184205	-2.588	0.0097 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4961 on 3974 degrees of freedom

Multiple R-squared: 0.04188, Adjusted R-squared: 0.04116

F-statistic: 57.9 on 3 and 3974 DF, p-value: < 2.2e-16

- ¿Qué significa la pendiente? Es la media esperada de la variable de respuesta, cuando la variable en cuestión toma una unidad de valor.
- En este caso, es la diferencia de peso entre los recién nacidos cuando donde "smoke == 0" versus "smoke == 1". En otras palabras, son las diferencias esperadas en peso de los recién nacidos para madres que no fuman, y las que si fuman. Esta diferencia es de .28 kg menos para los recién nacidos de madres que fuman, en contraste a las madres que no fuman.

Interacciones con una variable dicotómica: interacción

Residuals:

Min	1Q	Median	3Q	Max
-2.52762	-0.30262	0.00938	0.32138	1.79213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.477620	0.008508	408.741	<2e-16 ***
novisit	0.056733	0.120623	0.470	0.6381
smoke	-0.279750	0.022837	-12.250	<2e-16 ***
novisit:smoke	-0.476680	0.184205	-2.588	0.0097 **

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4961 on 3974 degrees of freedom

Multiple R-squared: 0.04188, Adjusted R-squared: 0.04116

F-statistic: 57.9 on 3 and 3974 DF, p-value: < 2.2e-16

- Los coeficientes de interacción pueden ser pensados como moderadores, los cuales pueden hacer dos cosas respecto a una relación de variables:
 - Pueden acelerar la relación
 - Pueden actuar como buffer de la relación anterior.
- Cuando una variable moderadora (i.e. una variable que interactúa con otra) y tiene un efecto de aceleración, comparte el mismo signo que la variable de interés. Y en términos sustantivos, lo que hace es “aumentar” el efecto de la variable de interés.
- Cuando una variable moderadora (i.e. una variable que interactúa con otra) y tiene un efecto de desaceleración, posee un signo contrario a la variable de interés. Y en términos sustantivos, lo que hace es “disminuir” el efecto de la variable de interés.

Interacciones con una variable dicotómica: interacción

JOHNSON-NEYMAN INTERVAL

When novisit is OUTSIDE the interval [-2.48, -0.32], the slope of smoke is $p < .05$.

Note: The range of observed values of novisit is [0.00, 1.00]

SIMPLE SLOPES ANALYSIS

Slope of smoke when novisit = 0.00 (0):

Est.	S.E.	t val.	p
-0.28	0.02	-12.25	0.00

Slope of smoke when novisit = 1.00 (1):

Est.	S.E.	t val.	p
-0.76	0.18	-4.14	0.00

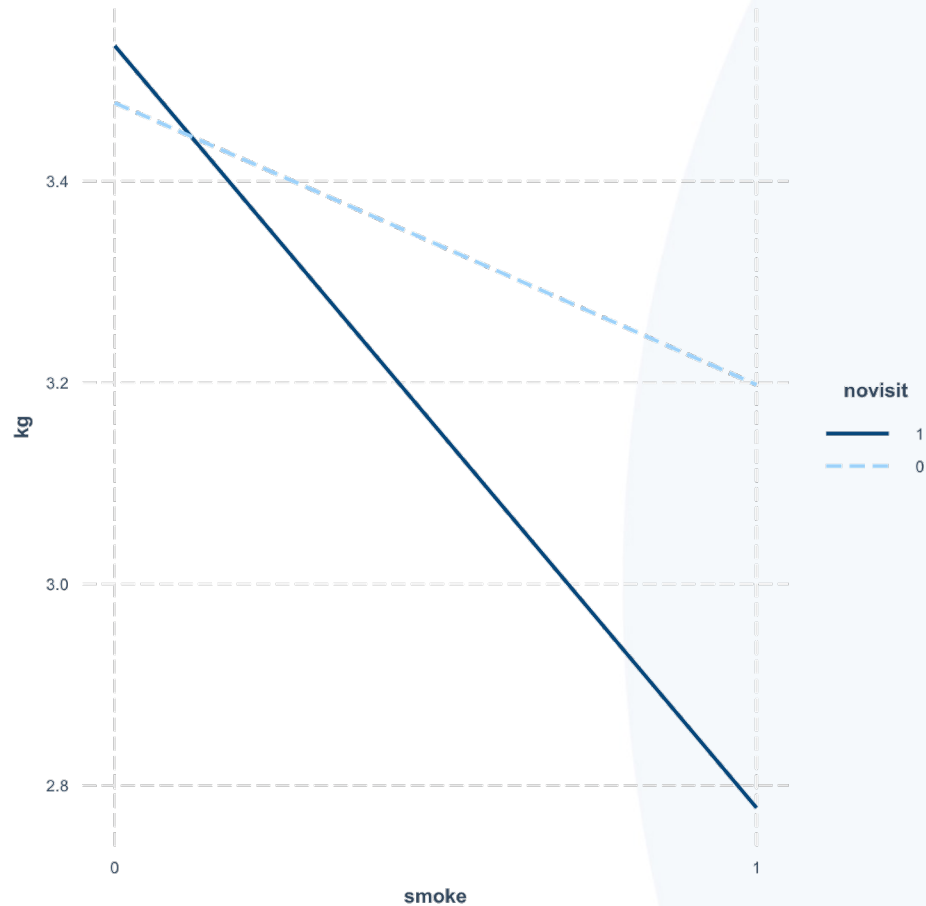
- En el caso de la presente variable tenemos un efecto de interacción que acelera la relación entre fumar, y el peso de los recién nacidos.
- Una forma de comunicar de forma más específica estos resultados esperados, es empleado los resultados de una prueba de "simple slopes".
- En este caso, indicamos que la diferencia de peso esperado de los recién nacidos, entre las madres que fuman, y las que no fuman es de -0.28 kg ($b = -0.28$, $SE = .02$, $p < .001$); pero esta diferencia esta moderada por la asistencia a los controles prenatales. Esta diferencia es aun mayor, para los nacidos de madres que fuman, y que no asisten a los controles pre-natales ($b = -0.76$, $SE = .18$, $p < .001$).
- El segundo término reportado, es básicamente la suma de $\beta + \delta$, es decir coeficiente que acompaña a smoke, mas el coeficiente que acompaña a la interacción.

Metodología Cuantitativa

Regresión con interacción

Notas sobre las interacciones

Notas sobre las interacciones



- Las interacciones como recurso de investigación, nos permiten responder a preguntas de tipo “cuándo”, “cómo”, o “bajo que condiciones”. Lo anterior, porque cuando una variable interactúa con otra, altera la relación entre dos variables condicionando al intercepto y a la pendiente que representa la relación de interés. En general, buscamos moderadores cuando queremos saber si una intervención funciona de forma similar para dos sub grupos, o cuando queremos saber si una relación es constante a diferentes condiciones en la población.
- Un aspecto importante a considerar es el poder estadístico requerido para evaluar una interacción. Si evaluar un intercepto y una pendiente requieren de 100 casos en una regresión, una interacción puede requerir cerca de 16 veces ese tamaño muestral (Gelman, 2018). Por tanto, incluir este tipo de términos en un modelo lineal requiere de ciertas cautelas.
- Un segundo aspecto a considerar es el centrado de variables. Centrar variables es crítico para hacer que los coeficientes del modelo sean interpretables, especialmente si se incluyen efectos de interacción. Varios componentes del modelo se pueden interpretar sin transformar variables; pero si queremos hacer que todo el modelo sea interpretable (e.g., interceptos y pendientes), es más conveniente realizar centrados de variables.

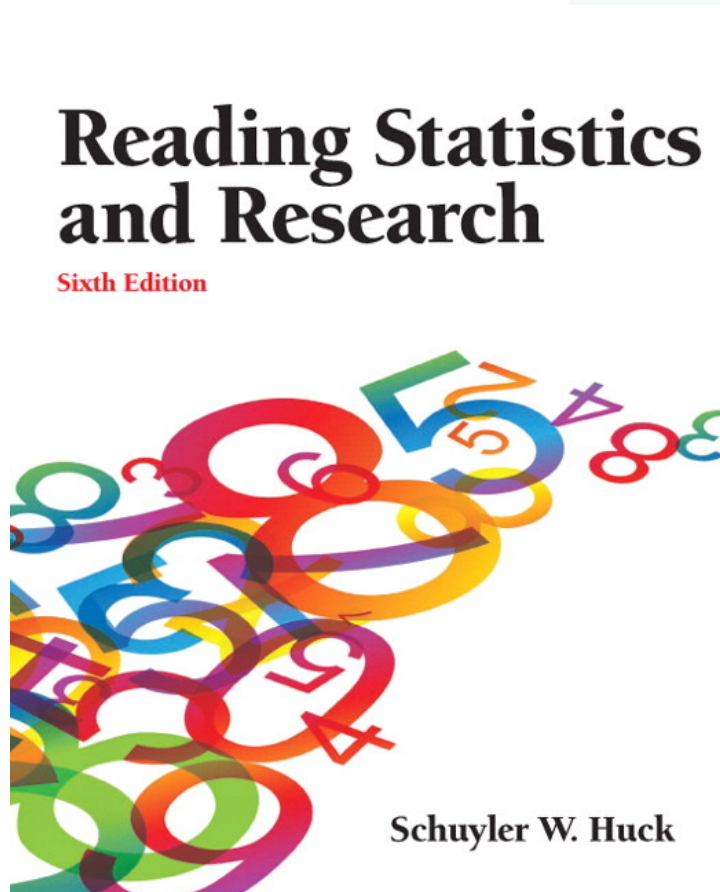
Metodología Cuantitativa

Referencias recomendadas

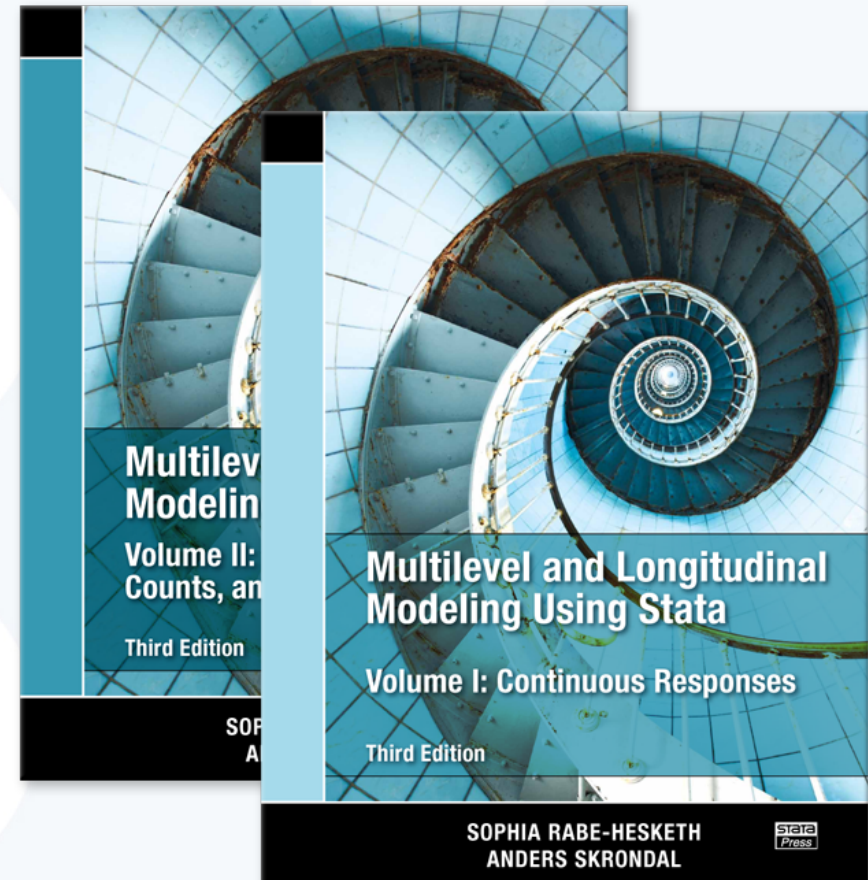
Libros de consulta

Referencias

Huck, S. W. (2012). *Reading Statistics and Research* (6th ed.). Pearson Education.



Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, Volumes I and II, Third Edition* (3rd ed.). Stata Press.



Metodología Cuantitativa

Regresión

Veamos cómo se ve esto en código!

Veamos esto en R!

https://github.com/dacarras/psi4035_examples/blob/master/psi4035_t06.rmd

Muchas gracias!

Referencias

Abrevaya, J. (2006). Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics*, 21(4), 489–519. <https://doi.org/10.1002/jae.851>

Gelman, A (2018) You need 16 times the sample size to estimate an interaction than to estimate a main effect Retrieved from: <https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/>

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, Volumes I and II, Third Edition* (3rd ed.). Stata Press.

Carrasco, D., PhD
Centro de Medición MIDE UC,
Pontificia Universidad Católica de Chile
<https://dacarras.github.io/>