



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

REGRESIÓN LOGÍSTICA MULTINIVEL

T E S I S

QUE PARA OBTENER EL TÍTULO DE
ACTUARIO

PRESENTA
CÉSAR BENJAMÍN LÓPEZ MORALES.

DIRECTOR DE TESIS:
MAT. MARGARITA ELVIRA CHÁVEZ CANO.



2013



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*A mis papás,
a mis hermanos,
a mis amigos.*

Agradecimientos

Agradezco en primer lugar a mi papá, Gerardo López, por su constante apoyo, por su consejo y quien siempre tendrá mi admiración.

A mi mamá, M. Concepción Morales, por su gran amor, comprensión, paciencia, por impulsarme a ser mejor y por guiarme siempre.

A mis hermanos, Ángel y Gerardo, por su cariño, por siempre creer en mi, por ser un excelente equipo.

A Jorge F. Tovar por su incondicional amistad, por estar siempre ahí. A Anaid Suárez, Alejandro Glass, Fernanda Lerdo de Tejada y Diego Valenzuela por su amistad, por tantas inolvidables experiencias.

A Eduardo S. Martínez por tantas risas y complicidades. A Karina Morales y Rocío Martínez por su gran amistad.

A Carlos Chida, Jesús Rodríguez, Orlando García, Mónica González, Karen Valencia, Karen R. Martínez, Daniel Armas, Cristina Collado y Paulina Barbosa por su confianza y porque siempre saben hacerme sonreír.

A la profesora Margarita Chávez Cano por su amistad, por su ejemplo, por sus consejos, por su confianza por brindarme la oportunidad de

trabajar junto a ella en sus cursos.

Al profesor Jaime Vázquez Alamilla por ser un gran ejemplo de excelencia y por su gran dedicación y entrega al desarrollo de la carrera de actuaría.

A la profesora Ruth Fuentes por su dedicación y sus acertadas sugerencias y consejos para enriquecer este trabajo.

A los profesores Arrigo Coen y Ángel Godoy por el tiempo dedicado a la revisión de este trabajo, apoyo y amistad.

Benjamín López Morales

Índice general

Introducción	1
1. Introducción al Modelo de Regresión Logística	5
1.1. Introducción	5
1.2. Planteamiento del modelo logístico	6
1.3. Ejemplo (Hosmer)	8
1.4. Comparativa entre la regresión logística y la regresión lineal múltiple	14
1.5. Pruebas de hipótesis de los parámetros de la regresión logística	16
1.6. Intervalos de confianza para los parámetros de la regresión logística	25
2. Análisis Multinivel	27
2.1. Introducción al análisis multinivel	27
2.2. Ejemplo	30
2.3. Coeficientes aleatorios	35
2.3.1. Estructuras de dos niveles	36
2.3.1.1. Modelo de intercepto aleatorio sin variables predictoras	36
2.3.1.2. Intercepto aleatorio con una variable predictora	39

2.3.1.3.	Intercepto y pendiente aleatorios con una variable predictora	41
2.3.1.4.	Intercepto y pendientes aleatorios con h variables predictoras	44
2.3.1.5.	Intercepto y pendiente aleatorios con una variable predictora en el primer nivel y una variable predictora al nivel de grupo	46
2.3.1.6.	Intercepto y pendiente aleatorios con h variables predictoras en el primer nivel y una variable predictora al nivel de grupo	47
2.3.1.7.	Intercepto y pendiente aleatorios con h variables predictoras en el primer nivel y L variables predictoras al nivel de grupo	47
2.3.2.	Estructuras con más de dos niveles	48
2.3.2.1.	Cuatro niveles sin variables predictoras	48
2.3.2.2.	Tres niveles con una variable predictora en cada nivel	49
2.4.	Estimación de parámetros	50
2.4.1.	El modelo de componentes de la varianza	50
2.4.2.	El modelo general de 2 niveles con coeficientes aleatorios	52
2.4.3.	Estimación por máxima verosimilitud utilizando mínimos cuadrados iterativos generalizados	53
3.	Regresión logística multinivel	57
3.1.	Modelo de intercepto aleatorio sin variables predictoras	57
3.1.1.	Ejemplo	58
3.2.	Modelo de intercepto aleatorio con una variable predictora	60
3.3.	Modelo de intercepto y pendiente aleatorios	60
3.3.1.	Ejemplo	61
3.4.	Estimación de parámetros de la regresión logística multinivel	68
3.5.	Proporciones como respuestas	68

Comentarios finales	71
Apéndices	73
A. Anexo al Capítulo 1	75
A.1. Cálculos previos al ajuste de los coeficientes	75
A.2. Ajuste de los coeficientes de regresión logística con el uso de R	77
B. Anexo al Capítulo 2	79
B.1. Ajuste de los coeficientes aleatorios con el uso de R	79
B.1.1. Modelo de intercepto aleatorio sin variables pre- dictoras	79
B.1.1.1. Modelo <i>complete-pooling</i>	80
B.1.1.2. Modelo <i>no-pooling</i>	80
B.1.1.3. Modelo multinivel	80
B.1.2. Intercepto aleatorio con una variable predictora .	82
B.1.2.1. Modelo <i>complete-pooling</i>	82
B.1.2.2. Modelo <i>no-pooling</i>	82
B.1.2.3. Modelo multinivel	83
B.1.3. Intercepto y pendiente aleatorios con una variable predictora	85
B.1.3.1. Modelo <i>complete-pooling</i>	85
B.1.3.2. Modelo <i>no-pooling</i>	86
B.1.3.3. Modelo multinivel	87
C. Producto Kronecker	89
Índice alfabético	95
Bibliografía	97

Introducción

El análisis multinivel es una metodología para el análisis de datos con patrones complejos de variabilidad. El enfoque principal radica en las fuentes de variación anidadas: alumnos dentro de salones de clases, empleados dentro de diversas compañías, mediciones repetidas dentro de los sujetos, individuos dentro de localidades, familias dentro de localidades dentro de los estados, pacientes dentro de clínicas dentro de hospitales.

En estudios a gran escala de salud pública, demográficos, y sociológicos, a menudo los datos siguen una estructura jerárquica debido a que las encuestas se basan en muestreo por conglomerados en varias etapas. Por lo tanto, el enfoque apropiado para el análisis de estos datos se basa en fuentes anidadas de variabilidad que provienen de los diferentes niveles y unidades de la jerarquía. Cuando la varianza de los errores residuales se relaciona con las observaciones individuales como resultado de estas estructuras anidadas, el modelo de regresión logística usual deja de ser apropiado.

En el análisis de los datos multinivel se pretende tomar en cuenta la variabilidad asociada a cada nivel de agregación. Existe variabilidad, por ejemplo, entre individuos pero también dentro de localidades y se podrían obtener conclusiones erróneas si cualquiera de estas fuentes de variabilidad se ignora.

En la investigación multinivel, la estructura de los datos en la población es jerárquica, y una muestra de dicha población puede ser vista como una muestra de múltiples etapas. Debido a consideraciones de costo, tiempo y eficiencia, las muestras estratificadas en varios niveles son la norma para los estudios sociológicos y demográficos. Para estas muestras el agrupamiento de los datos es un problema que debe ser tomado en cuenta en la fase de análisis de datos y la presentación de datos. Sin embargo, estas muestras, que son eficientes para el estudio descriptivo de la población, plantean muchos desafíos para la inferencia estadística basada en modelos usuales.

Este esquema de muestreo con agrupaciones a menudo introduce varios niveles de dependencia o correlación entre las observaciones que pueden tener implicaciones para las estimaciones de los parámetros del modelo. Para las muestras con múltiples niveles, la dependencia entre las observaciones a menudo proviene de varios niveles de la jerarquía. El problema de las dependencias entre las observaciones individuales también puede producirse en la investigación por encuestas, donde la muestra no se tomará al azar sino por medio de muestreo por conglomerados de zonas geográficas. En este caso, el uso de modelos estadísticos de un solo nivel ya no es válido ni razonable. Por lo tanto, con el fin de obtener inferencias y conclusiones pertinentes a partir de datos de encuestas estratificadas se requerirán técnicas de modelado con mayor complejidad como el análisis multinivel.

El análisis multinivel representa tanto las técnicas del análisis estadístico de esta clase de datos como la metodología para aplicarlos adecuadamente.

En el análisis de regresión logística multinivel se estudian muestras con estructura jerárquica en las que además las variables de respuesta tienen un rango dicotómico, es por eso que en el capítulo 1 se presenta

una introducción a la regresión logística. Este tipo de análisis servirá como base para el análisis de regresión logística multinivel ya que de manera natural se heredarán ciertas características.

En el capítulo 2 se presenta la teoría general del análisis multinivel para introducir el análisis general de datos con patrones complejos de variabilidad debido al anidamiento en clases de las observaciones.

Finalmente, en el capítulo 3 se estudiará el análisis logístico multinivel en el cual se manejan muestras que tienen una estructura jerárquica y en el cual las variables de respuesta son de rango dicotómico.

Capítulo 1

Introducción al Modelo de Regresión Logística

1.1. Introducción

El objetivo del modelo logístico de regresión, al igual que cualquier modelo de regresión es modelar con el mejor ajuste y de manera razonablemente sencilla la relación entre una o varias variables independientes (predictoras o explicativas) y una variable dependiente (o de respuesta). La particularidad del modelo de regresión logística es que se encarga del caso en que la variable de respuesta tiene rango dicotómico.

En el análisis de datos en ocasiones es de interés estudiar la relación de la probabilidad de éxito o fracaso en función de otros predictores. Es decir, si \mathbf{X} es una variable y Y una variable dependiente de \mathbf{X} , lo que se busca es encontrar la relación

$$P(Y = y|\mathbf{X} = \mathbf{x}),$$

para toda \mathbf{x} en el rango de \mathbf{X} y para $y = 0, 1$.

Se denota a $P(Y = 1|\mathbf{X} = \mathbf{x})$ como $p|\mathbf{x}$ o como $\pi(\mathbf{x})$.

1.2. Planteamiento del modelo logístico

Sea y_i el resultado de la i -ésima observación de una muestra aleatoria de tamaño n , donde $y_i = 1$ significa éxito y $y_i = 0$, fracaso. Es razonable pensar que estas observaciones son tales que $(Y_i|\mathbf{x}_i) \sim \text{Bernoulli}(\pi(\mathbf{x}_i))$, donde \mathbf{x}_i es el vector de variables predictoras de la i -ésima observación. Entonces, al ser las observaciones independientes, la función de verosimilitud es

$$\mathbf{L} = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}. \quad (1.1)$$

Por lo que la *log verosimilitud* es

$$\ell = \sum_{i=1}^n \left[y_i \log(\pi(\mathbf{x}_i)) + (1 - y_i) \log(1 - \pi(\mathbf{x}_i)) \right]. \quad (1.2)$$

Que puede ser reescrita como

$$\ell = \sum_{i=1}^n \left[y_i \log \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right) + \log(1 - \pi(\mathbf{x}_i)) \right].$$

Ahora, como la distribución Bernoulli pertenece a la familia exponencial, se puede expresar la densidad de $Y_i|\mathbf{x}_i$ de la forma

$$f_{Y_i|\mathbf{x}_i}(y; \theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]$$

Para funciones $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ tales que satisfacen

$$\mu = E(Y_i|\mathbf{x}_i) = \frac{\partial b(\theta)}{\partial \theta}$$

y

$$\text{Var}(Y_i|\mathbf{x}_i) = a(\phi) \frac{\partial^2 b(\theta)}{\partial \theta^2}.$$

La log verosimilitud de la muestra expresada como miembro de la familia exponencial es

$$\ell = \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right]. \quad (1.3)$$

De lo anterior se obtiene que $a(\phi) = \phi = 1$, $b(\theta) = \log(e^\theta + 1)$, $c(y, \phi) = 0$ y $\theta = \log\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right)$.

La liga canónica es una función g que relaciona una combinación afín de las variables predictoras con la esperanza de y_i , es decir, si $\mathbf{x}'_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ y si se define la transformación afín de los elementos de \mathbf{x}_i como

$$\eta = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{pi}. \quad (1.4)$$

Entonces la función liga g es tal que

$$g(E(Y_i | \mathbf{x}_i)) = \eta.$$

Y como $(Y_i | \mathbf{x}_i) \sim \text{Bernoulli}(\pi(\mathbf{x}_i))$, entonces

$$g(\pi(\mathbf{x}_i)) = \eta.$$

En este caso la liga canónica es θ , es decir,

$$\theta = g(\pi(\mathbf{x}_i)) = \log\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right).$$

Lo que lleva al *modelo lineal de regresión logística*

$$\log\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}. \quad (1.5)$$

De donde se obtiene la expresión de $\pi(\mathbf{x}_i)$ en términos de $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_p)$

$$\pi(\mathbf{x}_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}}. \quad (1.6)$$

La estimación del vector β se puede realizar por el método de máxima verosimilitud. Es decir, obteniendo los valores de β que maximicen la función de verosimilitud (1.1), o equivalentemente los valores que maximicen la función de log verosimilitud, a este vector resultante se le denomina *estimador máximo verosímil*.

1.3. Ejemplo (Hosmer)

ID	AGE	AGRP	CHD	ID	AGE	AGRP	CHD
1	20	1	0	30	36	3	0
2	23	1	0	31	37	3	0
3	24	1	0	32	37	3	1
4	25	1	0	33	37	3	0
5	25	1	1	34	38	3	0
6	26	1	0	35	38	3	0
7	26	1	0	36	39	3	0
8	28	1	0	37	39	3	1
9	28	1	0	38	40	4	0
10	29	1	0	39	40	4	1
11	30	2	0	40	41	4	0
12	30	2	0	41	41	4	0
13	30	2	0	42	42	4	0
14	30	2	0	43	42	4	0
15	30	2	0	44	42	4	0
16	30	2	1	45	42	4	1
17	32	2	0	46	43	4	0
18	32	2	0	47	43	4	0
19	33	2	0	48	43	4	1
20	33	2	0	49	44	4	0
21	34	2	0	50	44	4	0
22	34	2	0	51	44	4	1

Continúa en la siguiente página

ID	AGE	AGRP	CHD	ID	AGE	AGRP	CHD
23	34	2	1	52	44	4	1
24	34	2	0	53	45	5	0
25	34	2	0	54	45	5	1
26	35	3	0	55	46	5	0
27	35	3	0	56	46	5	1
28	36	3	0	57	47	5	0
29	36	3	1	58	47	5	0
59	47	5	1	80	57	7	0
60	48	5	0	81	57	7	0
61	48	5	1	82	57	7	1
62	48	5	1	83	57	7	1
63	49	5	0	84	57	7	1
64	49	5	0	85	57	7	1
65	49	5	1	86	58	7	0
66	50	6	0	87	58	7	1
67	50	6	1	88	58	7	1
68	51	6	0	89	59	7	1
69	52	6	0	90	59	7	1
70	52	6	1	91	60	8	0
71	53	6	1	92	60	8	1
72	53	6	1	93	61	8	1
73	54	6	1	94	62	8	1
74	55	7	0	95	62	8	1
75	55	7	1	96	63	8	1
76	55	7	1	97	64	8	0
77	56	7	1	98	64	8	1
78	56	7	1	99	65	8	1
79	56	7	1	100	69	8	1

Cuadro 1.1: *Edad y condición de presencia o ausencia de una condición cardiaca coronaria para 100 individuos.*

En el cuadro 1.1 se muestra la edad en años (AGE), y la existencia o ausencia de una enfermedad cardiaca coronaria (CHD) de 100 individuos seleccionados para participar en el estudio. Asimismo, se muestra una variable de distinción (ID) y una más de grupo de edad (AGRP). La variable de respuesta es CHD, que toma el valor de 0 para indicar la ausencia de CHD, o 1 para indicar que la enfermedad está presente en el individuo.

En este estudio de la población, es de interés explorar la relación entre la edad y la presencia o ausencia de CHD. Si la variable de respuesta fuera continua en lugar de binaria, posiblemente comenzaríamos el estudio haciendo un diagrama de dispersión de la variable dependiente contra la variable independiente. Este diagrama de dispersión se muestra en la figura 1.1. En este diagrama de dispersión todos los puntos caen en alguna de las dos líneas paralelas que representan la ausencia ($y = 0$) o presencia ($y = 1$) de CHD. Podemos observar que los individuos que no presentan evidencia de CHD tienden a ser más jóvenes que aquellos que sí la presentan. Aunque en la figura se aprecia claramente la naturaleza dicotómica de la variable de respuesta, no provee de una idea clara acerca de la relación entre las variables CHD y AGE. Un problema que aparece en la gráfica de la figura 1.1 es que la variabilidad de CHD es grande para todos los valores de AGE. Esto dificulta el describir la relación funcional entre la edad y la presencia de CHD. Un método que se puede utilizar para eliminar parte de la variación, manteniendo la estructura de la relación entre la variable de respuesta y la variable independiente, es crear intervalos para la variable independiente y calcular la media de la variable de respuesta en cada grupo. De esta manera se obtiene el cuadro 1.2.

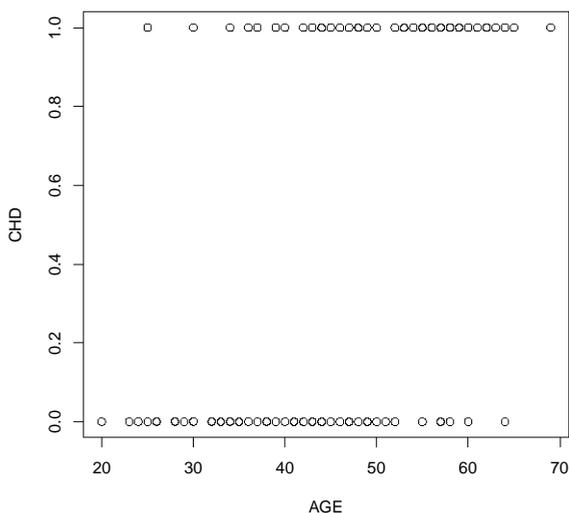


Figura 1.1: Diagrama de dispersión de CHD por AGE para 100 individuos.

Grupo de Edad	n	CHD		Media(Proporción)
		Ausente	Presente	
20-29	10	9	1	0.10
30-34	15	13	2	0.13
35-39	12	9	3	0.25
40-44	15	10	5	0.33
45-49	13	7	6	0.46
50-54	8	3	5	0.63
55-59	17	4	13	0.76
60-69	10	2	8	0.80
Total	100	57	43	0.43

Cuadro 1.2: Frecuencia de CHD por grupo de edad.

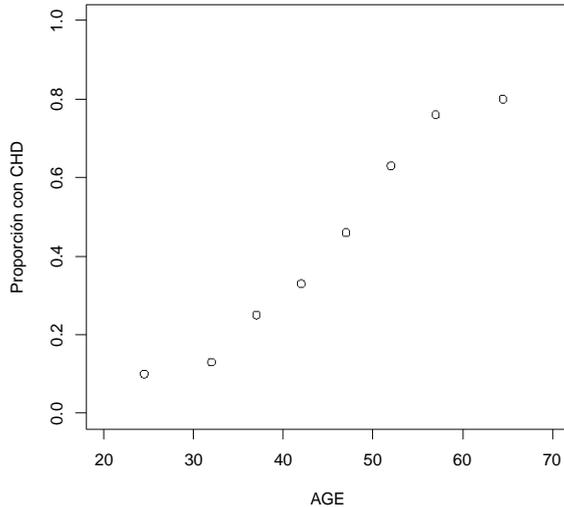


Figura 1.2: *Gráfica del porcentaje de individuos que presentan indicios de CHD por cada grupo de edad.*

Examinando este cuadro, una imagen más clara acerca de la relación comienza a aparecer. Aparentemente, a medida que la edad se incrementa, la proporción de individuos que presentan CHD aumenta. En la figura 1.2 se presenta una gráfica de la proporción de individuos que presentan CHD contra el punto medio de cada intervalo de edad. Esta gráfica muestra una imagen bastante clara de la relación entre la proporción de individuos que presentaron CHD y su edad, sin embargo, aún hace falta encontrar una relación funcional que la describa. Se pueden identificar 3 situaciones en la figura 1.2:

- Para edades grandes, se llega a un punto en el que la proporción de personas con indicios de CHD es tan alta que un aumento en la edad aporta poco a la proporción de personas con CHD.

- Una situación similar ocurre con edades pequeñas. En algún punto, la probabilidad de presentar CHD es tan baja que un pequeño cambio en la edad no afecta mucho la proporción de personas con CHD.
- Para personas con una edad media, digase alrededor de 50, se puede pensar que existe una relación lineal entre la edad y la proporción de personas con CHD.

Estos tres patrones sugieren que la curva de relación entre AGE y la probabilidad de presentar CHD tiene una forma de S. Como se plantea en la sección anterior, una curva con la forma adecuada es

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}.$$

Para estimar el vector de parámetros $\boldsymbol{\beta}' = (\beta_0, \beta_1)$ se deriva la función de log-verosimilitud $L(\boldsymbol{\beta})$ respecto a β_0 y β_1 y se igualan a cero. Así se obtienen las *ecuaciones de verosimilitud*

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (1.7)$$

y

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0. \quad (1.8)$$

En este caso, $n = 100$.

Estas ecuaciones no son lineales respecto a β_0 y β_1 , por lo que la solución del sistema requiere métodos especiales para su solución. En este trabajo se utiliza un método computacional para su solución numérica. Se denota $\hat{\boldsymbol{\beta}}$ al estimador máximo verosímil de $\boldsymbol{\beta}$. Al calcular los estimadores, se obtiene el cuadro 1.3.

Variable	Coefficiente	Desv. Est.	z	$P > z $
AGE	0.111	0.0241	4.61	< 0.001
Constante	-5.309	1.1337	-4.68	< 0.001
Log verosimilitud = -53.67656				

Cuadro 1.3: Resultados del ajuste de regresión logística para los datos del cuadro 1.1

Por lo que $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1) = (-5.309, 0.111)$, y por propiedades de los estimadores máximo verosímiles, se tiene que

$$\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = -5.309 + 0.111x$$

y

$$\hat{\pi}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x}} = \frac{e^{-5.309 + 0.111x}}{1 + e^{-5.309 + 0.111x}}.$$

1.4. Comparativa entre la regresión logística y la regresión lineal múltiple

El modelo de regresión lineal puede ser caracterizado de la siguiente manera. Sea \mathbf{U} una variable aleatoria con rango continuo y sea Z una variable dependiente de \mathbf{U} de rango continuo. Si se tienen observaciones $\{(\mathbf{u}_1, z_1), (\mathbf{u}_2, z_2), \dots, (\mathbf{u}_n, z_n)\}$ con $\mathbf{u}'_i = (u_{1i}, u_{2i}, \dots, u_{pi})$ con $i = 1, 2, \dots, n$, se supone que las observaciones satisfacen una relación lineal

$$z_i = \gamma_0 + \gamma_1 u_{1i} + \gamma_2 u_{2i} + \dots + \gamma_p u_{pi} + \varepsilon_i.$$

Donde el vector de coeficientes $\boldsymbol{\gamma}' = (\gamma_0, \gamma_1, \dots, \gamma_p)$ es desconocido y los errores ε_i se suponen con distribución $\varepsilon_i \sim N(0, \sigma^2)$, con σ^2 desconocida, de donde se tiene que

$$Z_i | \mathbf{u}_i \sim N(\gamma_0 + \gamma_1 u_{1i} + \gamma_2 u_{2i} + \dots + \gamma_p u_{pi}, \sigma^2).$$

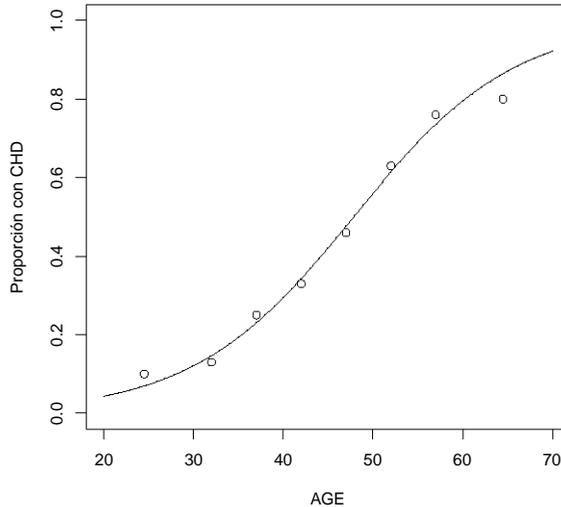


Figura 1.3: Gráfica del porcentaje observado de individuos que presentan indicios de CHD y la proporción estimada.

En el modelo de regresión logística, se tiene que \mathbf{X} es una variable aleatoria y que Y es una variable dependiente de \mathbf{X} de rango dicotómico. Si se tienen observaciones $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ con $\mathbf{x}'_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ con $i = 1, 2, \dots, n$, se supone que las observaciones se relacionan mediante la inversa de la función *logit*

$$y_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}}} + \varepsilon_i = \pi(\mathbf{x}_i) + \varepsilon_i.$$

Donde el vector de coeficientes $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_p)$ es desconocido y los errores ε_i pueden tener dos distintos valores. Si $y_i = 1$ entonces $\varepsilon_i = 1 - \pi(x)$ con probabilidad $\pi(x)$, y si $y = 0$, entonces $\varepsilon_i = -\pi(x)$ con probabilidad $1 - \pi(x)$. Por lo tanto ε_i se distribuye de tal manera

que tiene media cero y $Var(\varepsilon) = \pi(\mathbf{x}_i) [1 - \pi(\mathbf{x}_i)]$. Esto quiere decir que los errores pueden ser descritos por la distribución Bernoulli por lo que el modelo de regresión logística se basa en esta distribución, así

$$Y_i | \mathbf{x}_i \sim \text{Bernoulli}(\pi(\mathbf{x}_i))$$

1.5. Pruebas de hipótesis de los parámetros de la regresión logística

Una vez que se han estimado los coeficientes de predicción β es importante evaluar la significancia de las variables predictoras en el modelo. Esto involucra la formulación y pruebas de hipótesis para determinar si las variables independientes en el modelo son significativas con respecto a la variable de respuesta.

En el modelo logístico como en cualquier otro modelo de regresión surgen en este punto las preguntas:

- ¿Las variables independientes proporcionan bajo este modelo una correcta predicción del resultado de la variable de respuesta?
- ¿Algunas de las variables proveen mayor poder de predicción, mientras que otras aportan poco o nada al modelo?

Es decir, se quiere averiguar si la predicción de la variable de respuesta con ciertas variables predictoras en el modelo es mejor cuando éstas se encuentran en el modelo, que cuando se excluyen al realizar el análisis, si esto ocurre, entonces se dice que dichas variables son *significativas*. La manera general de evaluar la significancia de las variables es claramente ilustrada en el modelo de regresión lineal. En este modelo, la evaluación de la significancia de los coeficientes se realiza a través de la *tabla de análisis de varianza*. Esta tabla particiona la suma de las diferencias cuadradas de las observaciones con respecto a sus valores medios (SST) en la suma de las diferencias cuadradas de las observaciones con respecto a la línea de regresión (SSE), (o suma de los cuadrados de los residuales),

y la suma de los cuadrados de los valores esperados con respecto a la media de la variable dependiente (SSR), es decir,

$$SSR = \sum_{i=1}^n (\hat{z}_i - \bar{z}_i)^2,$$

$$SSE = \sum_{i=1}^n (z_i - \hat{z}_i)^2,$$

$$SST = \sum_{i=1}^n (z_i - \bar{z}_i)^2 = SSR + SSE.$$

Donde z_i es el valor de la i -ésima observación, \hat{z}_i es la estimación de la observación bajo el ajuste y \bar{z} es la media de las observaciones.

La tabla de análisis de varianza se muestra en el cuadro 1.4.

Tabla de Análisis de varianza ($H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_m = 0$)

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F
Regresión	$p - 1$	SSR	$SSR/(p - 1)$	$\frac{SSR}{p-1}$
Residual	$n - p$	SSE	$SSE/(n - p)$	$\frac{SSE}{n-p}$
Total	$n - 1$	SST		

Cuadro 1.4: *Tabla de análisis de varianza para la regresión lineal*

Entonces para probar si algún coeficiente proporciona poder de predicción para la variable de respuesta, es decir una prueba general de la significancia de la regresión, prueba

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$$

contra

$$H_a : \text{algún } \gamma_j \neq 0, \quad j = 1, 2, \dots, p.$$

La estadística de prueba F de la última columna de la tabla de análisis de varianza se compara contra una distribución F con $(p, n - p)$ grados

de libertad.

Para responder la segunda pregunta, es decir, para probar el poder de predicción de cada variable explicativa, se prueba la significancia del coeficiente correspondiente, es decir,

$$H_0 : \gamma_j = 0, \quad j = 1, 2, \dots, p$$

contra

$$H_a : \gamma_j \neq 0.$$

Para esto se utiliza una estadística de prueba t

$$t_j = \frac{\hat{\gamma}_j}{\widehat{S.E.}(\hat{\gamma}_j)}, \quad (1.9)$$

que es comparado con una distribución t con $n-p-1$ grados de libertad. Los valores de $\widehat{S.E.}(\hat{\gamma}_j)$ se obtienen como resultado de las raíces cuadradas de los elementos de la diagonal de la matriz $\widehat{Var}(\hat{\gamma}) = (\mathbf{u}'\mathbf{u})^{-1} \hat{\sigma}^2$, donde $\hat{\sigma}^2$ es el cuadrado medio de los residuales

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{n-p}.$$

Y la matriz \mathbf{u} , definida como:

$$\mathbf{u} = \begin{pmatrix} 1 & u_{11} & u_{12} & \cdots & u_{1p} \\ 1 & u_{21} & u_{22} & \cdots & u_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_{n1} & u_{n2} & \cdots & u_{np} \end{pmatrix}.$$

La prueba F y la prueba t , las herramientas en la inferencia de mínimos cuadrados, son generalizadas en el cociente de verosimilitudes y en la estadística de Wald en regresión logística. El objetivo principal de las pruebas de hipótesis en regresión logística es el mismo que en regresión lineal: comparar si el modelo es más adecuado cuando se incluyen ciertas

variables de predicción que cuando estas se excluyen del modelo. Estas comparaciones entre valores observados y esperados, en regresión logística se basan en la función de log verosimilitud.

La estadística dada por el cociente de verosimilitudes compara la hipótesis nula de un modelo saturado (más simple) contra la hipótesis alternativa de un modelo más general. Un modelo saturado es aquel que incluye el mismo número de observaciones que parámetros en el modelo (un ejemplo simple de modelo saturado es ajustar un modelo de regresión lineal con únicamente dos datos, $n = 2$). El cociente de verosimilitudes se expresa de la siguiente manera

$$LR = \frac{(\text{verosimilitud del modelo ajustado})}{(\text{verosimilitud del modelo saturado})}. \quad (1.10)$$

En un modelo saturado, por definición del modelo mismo, se tiene que $E(y_i|x) = y_i$, así, en el modelo logístico se tiene que en un modelo saturado $\hat{\pi}(\mathbf{x}) = y_i$, por lo que la verosimilitud será

$$(\text{verosimilitud del modelo saturado}) = \prod_{i=1}^p y_i^{y_i} \times (1 - y_i)^{(1-y_i)} = 1. \quad (1.11)$$

La comparación de los valores observados con los valores esperados utilizando la función de verosimilitud está basada en la estadística D llamada *devianza*, definida como

$$D = -2 \log(LR) = -2 \log \left[\frac{(\text{verosimilitud del modelo ajustado})}{(\text{verosimilitud del modelo saturado})} \right]. \quad (1.12)$$

La devianza es equivalente a la suma de cuadrados de residuales en la regresión lineal. De hecho, la devianza, como se expresa en la ecuación (1.12), al calcularse para la regresión lineal, es idéntica a SSE . Entonces, por las expresiones (1.11) y (1.12), se tiene que

$$D = -2 \log(\text{verosimilitud del modelo ajustado}).$$

Para evaluar la significancia de las variables de predicción, se compara el valor de D obtenido al incluir la variable en el modelo y el valor obtenido al excluirla, es decir,

$$G = D(\text{modelo excluyendo la variable}) - D(\text{modelo incluyendo la variable}).$$

equivalentemente,

$$G = -2 \log \left[\frac{(\text{verosimilitud excluyendo las variables})}{(\text{verosimilitud incluyendo las variables})} \right] \quad (1.13)$$

Para calcular la verosimilitud del modelo excluyendo las variables explicativas, se tiene que, al suponer $\beta_j = 0$ para $j = 1, 2, \dots, m$, la probabilidad de éxito de la variable de respuesta, denotada $\hat{\pi}^0$, estará dada por

$$\hat{\pi}^0 = \frac{e^{\hat{\beta}_0}}{1 + e^{\hat{\beta}_0}}. \quad (1.14)$$

Por lo que hay que encontrar el estimador máximo verosímil de β_0 bajo este modelo. Por la ecuación (1.2) se tiene que la log verosimilitud del modelo excluyendo las variables explicativas es

$$L(\beta_0) = \sum_{i=1}^n [y_i \log(\pi^0) - ((1 + y_i) \log(1 - \beta_0))]. \quad (1.15)$$

Entonces, por la ecuación (1.7), se tiene que

$$\frac{\partial L(\beta_0)}{\partial \beta_0} = \sum_{i=1}^n [y_i - \hat{\pi}^0] = 0. \quad (1.16)$$

Por lo que el estimador máximo verosímil de π^0 es

$$\hat{\pi}^0 = \frac{\sum_{i=1}^n y_i}{n}. \quad (1.17)$$

Si se denota

$$n_1 = \sum_{i=1}^n y_i \quad \text{y} \quad n_0 = \sum_{i=1}^n (1 - y_i), \quad (1.18)$$

entonces,

$$\hat{\pi}^0 = \frac{n_1}{n} \quad \text{y} \quad (1 - \hat{\pi}^0) = \frac{n_0}{n}. \quad (1.19)$$

Aplicando la función logit expresada en la ecuación (1.5) y por propiedades de los estimadores máximo verosímiles

$$\log \left(\frac{\hat{\pi}^0}{1 - \hat{\pi}^0} \right) = \hat{\beta}_0.$$

Por lo que el estimador máximo verosímil de β_0 , bajo el modelo excluyendo las variables predictoras es

$$\hat{\beta}_0 = \log \left(\frac{n_1}{n_0} \right).$$

De lo anterior, se tiene que la verosimilitud del modelo excluyendo las variables independientes es

$$(\text{verosimilitud excluyendo las variables}) = \prod_{i=1}^n \left[\left(\frac{n_1}{n} \right)^{y_i} \left(\frac{n_0}{n} \right)^{(1-y_i)} \right].$$

De donde se obtiene que

$$(\text{verosimilitud excluyendo las variables}) = \left(\frac{n_1}{n} \right)^{n_1} \left(\frac{n_0}{n} \right)^{n_0}. \quad (1.20)$$

Por lo tanto, de las ecuaciones (1.1), (1.20) y (1.13), se tiene que

$$G = -2 \log \left[\frac{\left(\frac{n_1}{n} \right)^{n_1} \left(\frac{n_0}{n} \right)^{n_0}}{\prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}} \right]. \quad (1.21)$$

Entonces, para probar

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

contra

$$H_a : \text{algún } \beta_j \neq 0, \quad j = 1, 2, \dots, p,$$

se utiliza la estadística G que tiene una distribución ji cuadrada con p grados de libertad, se denota $G \sim \chi_p^2$. Esto es, para un nivel de significancia α , se rechazará la hipótesis nula si $G > \chi_p^2 (1-\alpha)$.

Para hablar de la siguiente estadística de prueba, hace falta primero mencionar el cálculo de las varianzas y covarianzas de los parámetros de la regresión. Este cálculo se realiza por el método de máxima verosimilitud. Este método establece que los estimadores de los coeficientes se obtienen de las segundas derivadas parciales de la función de log verosimilitud. Éstas tienen la siguiente forma:

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i)$$

y

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n x_{ij} x_{ik} \pi_i (1 - \pi_i)$$

para $j, k = 0, 1, 2, \dots, p$, donde π_i denota $\pi(\mathbf{x}_i)$. Sea $\mathbf{I}(\boldsymbol{\beta})_{jk} = -\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}$. Esta matriz es llamada la *matriz de información observada*. Las varianzas y covarianzas se obtienen de la inversa de esta matriz, por lo que se denota $\text{Var}(\boldsymbol{\beta}) = \mathbf{I}^{-1}(\boldsymbol{\beta})$. Sin embargo, no hay una manera analítica de expresar esta matriz salvo en casos especiales, por lo cual se usa la notación $\text{Var}(\beta_j)$ para referirse al j -ésimo elemento diagonal de esta matriz y la notación $\text{Cov}(\beta_j, \beta_k)$, para denotar a $\mathbf{I}^{-1}(\boldsymbol{\beta})_{jk}$ con $j \neq k$. Los estimadores máximo verosímiles de las varianzas y covarianzas son obtenidos, por propiedades de los estimadores máximo verosímiles, evaluando $\text{Var}(\boldsymbol{\beta})$ en $\hat{\boldsymbol{\beta}}$, a esta matriz se le denota $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})$ y al j -ésimo elemento diagonal de esta matriz se le denomina $\widehat{\text{Var}}(\hat{\beta}_j)$, y al elemento $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})_{jkl}$ se le denomina $\widehat{\text{Cov}}(\hat{\beta}_j, \hat{\beta}_k)$ para $j, k = 0, 1, 2, \dots, p$. Entonces

se define el error estándar estimado de $\hat{\beta}_j$ como

$$\widehat{s.e.}(\hat{\beta}_j) = \left(\widehat{Var}(\hat{\beta}_j)\right)^{\frac{1}{2}},$$

para $j = 0, 1, 2, \dots, p$.

Una manera de calcular la matriz de información observada, es de la siguiente manera

$$\widehat{\mathbf{I}}(\hat{\boldsymbol{\beta}}) = \mathcal{X}'\mathcal{V}\mathcal{X}$$

donde \mathcal{X} es la matriz que tiene la información de cada observación y \mathcal{V} es una matriz diagonal, cuyo j -ésimo elemento diagonal es de la forma $\hat{\pi}_i(1 - \hat{\pi}_i)$. Así, la matriz \mathcal{X} es

$$\mathcal{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

y la matriz \mathcal{V}

$$\mathcal{V} = \begin{pmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & 0 & \cdots & 0 & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & 0 & \cdots & 0 & 0 \\ 0 & 0 & \hat{\pi}_3(1 - \hat{\pi}_3) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \hat{\pi}_{n-1}(1 - \hat{\pi}_{n-1}) & 0 \\ 0 & 0 & 0 & \cdots & 0 & \hat{\pi}_n(1 - \hat{\pi}_n) \end{pmatrix}.$$

Así, para probar la significancia de un coeficiente de regresión individual, β_j con $j = 1, 2, \dots, p$, es decir, para probar

$$H_0 : \beta_j = 0$$

contra

$$H_a : \beta_j \neq 0$$

se utiliza la prueba de Wald univariada, la estadística de prueba se expresa de la siguiente manera,

$$W_j = \frac{\widehat{\beta}_j}{\widehat{s.e.}(\widehat{\beta}_j)}$$

Bajo la hipótesis nula, esta estadística se distribuye normal estándar, por lo que se rechazará H_0 si $\left[W_j > z_{(1-\frac{\alpha}{2})}\right]$ ó $\left[W_j < z_{(\frac{\alpha}{2})}\right]$.

Donde $z_{(1-\frac{\alpha}{2})}$ y $z_{(\frac{\alpha}{2})}$ son los cuantiles $1 - \frac{\alpha}{2}$ y $\frac{\alpha}{2}$ respectivamente de una distribución normal estándar.

La prueba de Wald univariada se generaliza para el caso en que se prueba la hipótesis más general

$$H_0 : \beta_j = \beta_{j0}$$

contra

$$H_a : \beta_j \neq \beta_{j0}$$

para cualquier β_{j0} y para $j = 1, 2, \dots, p$, con la estadística de prueba

$$W_{j,\beta_{j0}} = \frac{\widehat{\beta}_j - \beta_{j0}}{\widehat{s.e.}(\widehat{\beta}_j)}$$

Por lo que para un nivel de significancia α , se rechazará H_0 si $\left[W_{j,\beta_{j0}} > z_{(1-\frac{\alpha}{2})}\right]$ ó $\left[W_{j,\beta_{j0}} < z_{(\frac{\alpha}{2})}\right]$.

Existe una prueba análoga a la de Wald para el caso en que se quiera probar la significancia de las variables explicativas, es decir, si se quiere probar

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_p$$

contra

$$H_a : \text{algún } \beta_j \neq 0, \quad j = 0, 1, 2, \dots, p$$

entonces se utiliza la estadística

$$W = \widehat{\beta}' \left[\widehat{Var}(\widehat{\beta}) \right]^{-1} \widehat{\beta}$$

que se distribuye χ_{p+1}^2 bajo la hipótesis nula. Para probar solamente la significancia de los coeficientes de las variables predictoras, se elimina $\widehat{\beta}_0$ de $\widehat{\beta}$, y la columna asociada a $\widehat{\beta}_0$ de \mathcal{X} .

De manera similar, se puede probar

$$H_0 : \beta_{j_1} = \beta_{j_2} = \cdots = \beta_{j_s} = 0$$

contra

$$H_a : \text{algún } \beta_{j_i} \neq 0, \quad j_i = j_1, j_2, \cdots, j_s$$

donde $j = \{j_1, j_2, \cdots, j_s\}$ es un subconjunto de índices de $\{0, 1, \cdots, p\}$.

1.6. Intervalos de confianza para los parámetros de la regresión logística

Los mismos principios estadísticos que se utilizan para probar la significancia de los parámetros del modelo se utilizan en la construcción de los intervalos de confianza. Para encontrar los intervalos de confianza de los coeficientes de las variables de predicción así como del término constante se utilizan sus respectivas pruebas de Wald univariadas. Así, el intervalo de confianza de $100(1 - \alpha)\%$ para el parámetro β_j con $j = 0, 1, 2, \cdots, p$, será

$$\widehat{\beta}_j \pm z_{(1-\frac{\alpha}{2})} \widehat{s.e.}(\widehat{\beta}_j)$$

donde $z_{(1-\frac{\alpha}{2})}$ es el cuantil $(1 - \frac{\alpha}{2})$ de la distribución normal estándar. Para calcular el intervalo de confianza para el modelo lineal de regresión logística, primero se definirá lo siguiente,

$${}_1\mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \text{ donde } x_0 = 1.$$

Así se tiene que

$$\widehat{g}_\pi(\mathbf{1}\mathbf{x}) = \mathbf{1}\mathbf{x}'\widehat{\boldsymbol{\beta}} = \widehat{\beta}_0 + \widehat{\beta}_1x_1 + \widehat{\beta}_2x_2 + \cdots + \widehat{\beta}_px_p.$$

Para calcular el intervalo de confianza del modelo lineal de regresión logística evaluado en $\mathbf{1}\mathbf{x}$, primero se estima su varianza, es decir,

$$Var(g_\pi(\mathbf{1}\mathbf{x})) = Var\left(\widehat{\beta}_0 + \widehat{\beta}_1x_1 + \widehat{\beta}_2x_2 + \cdots + \widehat{\beta}_px_p\right).$$

que, por propiedades de la varianza de la suma de variables aleatorias, se tiene que

$$Var(g_\pi(\mathbf{1}\mathbf{x})) = \sum_{j=0}^m x_j^2 Var(\widehat{\beta}_j) + \sum_{j=0}^p \sum_{k=j+1}^p 2x_jx_k Cov(\widehat{\beta}_j, \widehat{\beta}_k)$$

Que por propiedades de los estimadores máximo verosímiles se tiene

$$\widehat{Var}(\widehat{g}_\pi(\mathbf{1}\mathbf{x})) = \sum_{j=0}^m x_j^2 \widehat{Var}(\widehat{\beta}_j) + \sum_{j=0}^p \sum_{k=j+1}^p 2x_jx_k \widehat{Cov}(\widehat{\beta}_j, \widehat{\beta}_k).$$

Expresando lo anterior de manera matricial

$$\widehat{Var}(\widehat{g}_\pi(\mathbf{1}\mathbf{x})) = \mathbf{1}\mathbf{x}'\widehat{Var}(\widehat{\boldsymbol{\beta}})\mathbf{1}\mathbf{x} = \mathbf{1}\mathbf{x}'(\mathcal{X}'\mathcal{V}\mathcal{X})^{-1}\mathbf{1}\mathbf{x}.$$

Entonces, se define

$$\widehat{s.e.}(\widehat{g}_\pi(\mathbf{1}\mathbf{x})) = \left(\widehat{Var}(\widehat{g}_\pi(\mathbf{1}\mathbf{x}))\right)^{\frac{1}{2}}.$$

Por lo tanto, los extremos del intervalo de confianza de $100(1 - \alpha)\%$ para $g_\pi(\mathbf{1}\mathbf{x})$ son de la forma

$$\widehat{g}_\pi(\mathbf{1}\mathbf{x}) \pm z_{(1-\frac{\alpha}{2})}\widehat{s.e.}(\widehat{g}_\pi(\mathbf{1}\mathbf{x})).$$

Cuando no hay lugar a confusión, se usa la notación \mathbf{x} para referirse a $\mathbf{1}\mathbf{x}$.

Capítulo 2

Análisis Multinivel

2.1. Introducción al análisis multinivel

La investigación social a menudo se refiere a la relación entre los individuos y los grupos a los que éstos pertenecen. Esto conduce a modelos jerárquicos o datos con estructuras multinivel, en los cuales los datos están anidados dentro de grupos.

Los datos que tienen una estructura anidada o jerárquica surgen en una gran variedad de campos de estudio, por ejemplo en la investigación educativa, en los cuales los alumnos están agrupados en escuelas; en estudios sobre la familia, en los cuales los niños son agrupados dentro de familias; en la investigación médica en los cuales los pacientes son agrupados en médicos tratantes o en hospitales; y en investigación biomédica, por ejemplo, el análisis de anomalías dentales en los cuales las piezas dentales están agrupadas dentro de individuos. Los datos agrupados también pueden surgir como resultado del diseño específico de la investigación. Por ejemplo, en un estudio de investigación a gran escala, la recolección de los datos es usualmente organizada en un diseño de muestreo multietapa que resulta en datos conglomerados o estratificados. Otro ejemplo es el diseño longitudinal en los cuales los datos son

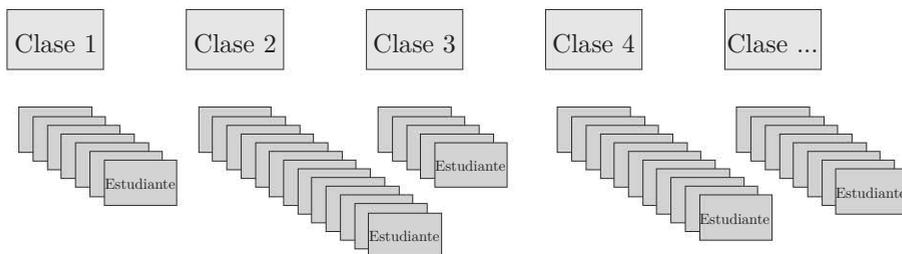


Figura 2.1: *Estructura de datos de dos niveles.*

una serie de mediciones repetidas agrupadas en sujetos individuales. Un problema importante en el análisis estadístico de datos con estructura jerárquica es la dependencia de las observaciones en los niveles más bajos o profundos. Enfoques más antiguos al análisis multinivel simplemente ignoran este problema y comúnmente realizan el análisis desagregando todos los datos al nivel más bajo y aplicando métodos estándar de análisis.

El análisis multinivel fue desarrollado en un principio para la investigación educativa. Analizando el desempeño de los estudiantes, los investigadores se dieron cuenta de que las observaciones de estudiantes que pertenecían a una misma clase no eran independientes entre sí. Como las herramientas estadísticas *estándar* tienen el supuesto que las observaciones son independientes, entonces no es apropiado utilizar estas técnicas para analizar estos datos. Los estudiantes que pertenecen a una misma clase pueden ser pensados como un tipo de jerarquía; los alumnos están agrupados en clases. Esto es conocido como una estructura de datos de dos niveles, el primer nivel serían los estudiantes y el segundo, las clases. Esto se ilustra en la figura 2.1.

La idea general del análisis multinivel es que esta jerarquía se tome en cuenta en el análisis, en otras palabras, tomar en cuenta la relación que hay entre las observaciones.

En esta estructura de datos se puede observar que quizá no solamente se pueden considerar a los alumnos agrupados en clases, sino que además

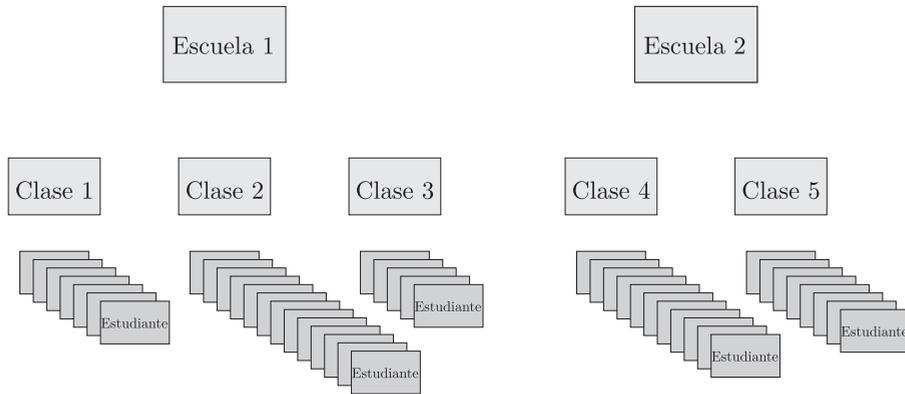


Figura 2.2: *Estructura de datos de tres niveles.*

estas últimas se podrían considerar agrupadas en escuelas. A esto se le conoce como una estructura de datos de tres niveles, donde el nivel más profundo, el nivel uno, son los estudiantes, el nivel dos son las clases y el tres, las escuelas.

La idea detrás del análisis multinivel es tomar en cuenta la dependencia que existe entre las observaciones, no solo entre clases sino además entre escuelas.

En el análisis multinivel, las variables dependientes pueden estar definidas en cualquier nivel de la jerarquía. Algunas de estas variables se pueden medir al nivel en que están definidas, por ejemplo en el nivel "Escuela" del ejemplo anterior, se puede medir, por ejemplo, el tamaño de la escuela o su denominación, y al nivel "Estudiante" se puede medir la inteligencia o el desempeño escolar. Además, se pueden mover las variables de un nivel a otro por *agregación* o por *desagregación*. Agregación significa que una variable definida en un nivel es movida a un nivel superior, por ejemplo, asignando a cada escuela la media del puntaje de inteligencia de sus alumnos. Desagregación significa mover una variable a un nivel inferior, por ejemplo asignando a los alumnos la denominación de la escuela a la que pertenecen.

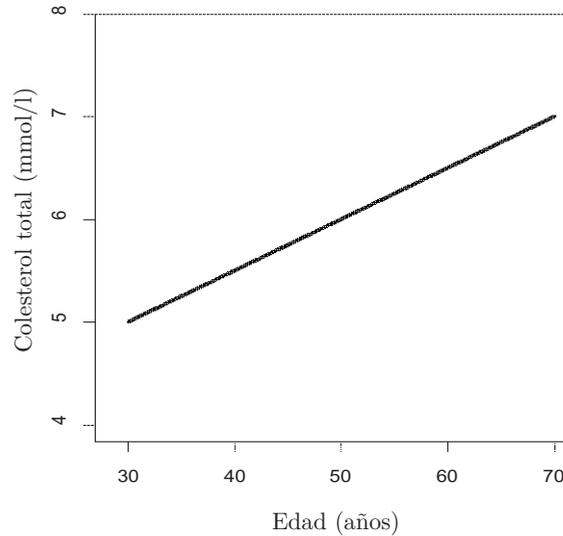


Figura 2.3: Análisis de regresión lineal de la relación entre la edad y el colesterol total.

2.2. Ejemplo

Supóngase que se tiene un estudio para investigar la relación que existe entre el colesterol total y la edad. La figura 2.3 muestra los resultados del análisis de regresión lineal y la siguiente ecuación describe el modelo de regresión lineal

$$\text{Colesterol total} = \beta_0 + \beta_1 \times \text{edad} + \varepsilon.$$

Supóngase ahora que se quiere corregir el análisis por género. Y como el comportamiento del colesterol en relación con la edad es distinto en hombres y en mujeres, se tendrá

$$\text{Colesterol total} = \beta_0 + \beta_1 \times \text{edad} + \beta_2 \times \text{sexo} + \varepsilon,$$

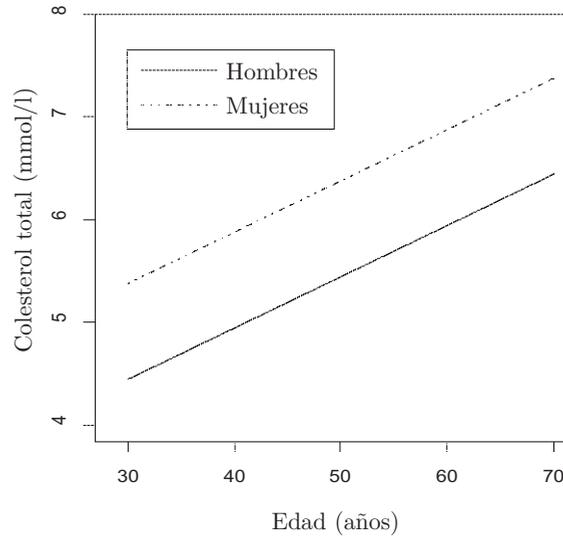


Figura 2.4: *Análisis de regresión lineal de la relación entre la edad y el colesterol total corregido por género.*

donde β_2 es el coeficiente de regresión por sexo.

Supóngase que los hombres son codificados con 0 y las mujeres con 1, entonces la β_0 es el término constante para hombres y $\beta_0 + \beta_2$ es la constante para mujeres. Así que una corrección por género significa que el término constante se supone distinto para hombres y para mujeres, como se muestra en la figura 2.4.

Por otro lado, si además los pacientes están agrupados respecto al médico que los atiende, se puede pensar que los pacientes que ‘*pertenecen*’ a un médico tienen características que difieren de aquellos pacientes que ‘*pertenecen*’ a otro médico. Es por esto que se quiere hacer una corrección por médico en el análisis de regresión. De nuevo, y como se observa en la figura 2.5, esto significa que se estimarán los términos constantes de la regresión para cada médico.

Sin embargo, al hacer esta corrección surge el problema de que la varia-

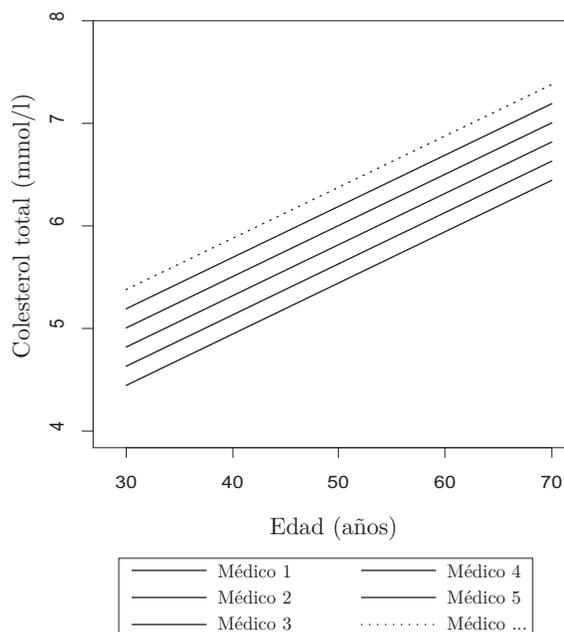


Figura 2.5: Análisis de regresión lineal de la relación entre la edad y el colesterol total corregido por médico.

ble ‘médico’ no es continua y además no es dicotómica. Esta variable es categórica (más específicamente, nominal), cuando esto sucede, es necesario crear variables *dummy* o variables *mudas*, dependiendo del número de médicos involucrados en el estudio (que será el número de médicos menos uno), y para todas esas variables dummy se deben calcular los coeficientes de regresión correspondientes, es decir, para m médicos involucrados en el estudio, se tendrá que

$$\begin{aligned} \text{Colesterol total} = & \beta_0 + \beta_1 \text{edad} + \beta_2 \text{dummyMD}_1 + \beta_3 \text{dummyMD}_2 + \\ & \dots + \beta_m \times \text{dummyMD}_{m-1} + \varepsilon \end{aligned}$$

donde desde β_2 hasta β_m son los coeficientes de regresión que representan a los distintos médicos.

Así, si hay 12 médicos en el estudio, se deben estimar 11 coeficientes de regresión adicionales. Esto es muy poco eficiente, dado que la variable ‘médico’ se añadió al modelo para hacer una corrección del mismo, y no hay un interés real en conocer los distintos niveles de colesterol para los pacientes pertenecientes a cada médico. Una manera de corregir por médico está dada por el análisis multinivel. Utilizando este análisis no se calculan cada uno de los términos constantes de regresión, sino que se estima la varianza de ellos, es decir, se supone que estos son aleatorios. En general, el análisis multinivel es una manera eficiente de corregir con respecto a una variable categórica que puede tomar muchos valores. Sin embargo, hay una especie de intercambio ya que en el análisis multinivel se hace la suposición de que estos parámetros aleatorios tienen una distribución normal.

Regresando al ejemplo donde se analizó la relación entre edad y colesterol total con una corrección por género, supóngase ahora que además de suponer que los términos constantes son distintos (figura 2.6), también se considera que la pendiente de la relación es distinta para cada sexo, para esto, se tiene que añadir al modelo un coeficiente para la interacción entre género y edad, así, se estiman distintas pendientes para la línea de regresión para hombres y para mujeres, es decir,

$$\text{Colesterol total} = \beta_0 + \beta_1 \text{ edad} + \beta_2 \text{ sexo} + \beta_3(\text{edad} \times \text{sexo}) + \varepsilon$$

donde β_3 es el coeficiente de regresión para la interacción entre edad y sexo.

Cuando esta variable categórica es dicotómica, sólo se añade un término de interacción al modelo de regresión, sin embargo, cuando esta variable toma muchos distintos valores, por ejemplo, cuando se quiere hacer una corrección por médico, surge el problema que se describió anteriormente. Así, es razonable suponer que la relación entre colesterol total y la edad es distinta para los pacientes pertenecientes a distintos médicos. Es decir, Se tendrán que calcular distintas pendientes para

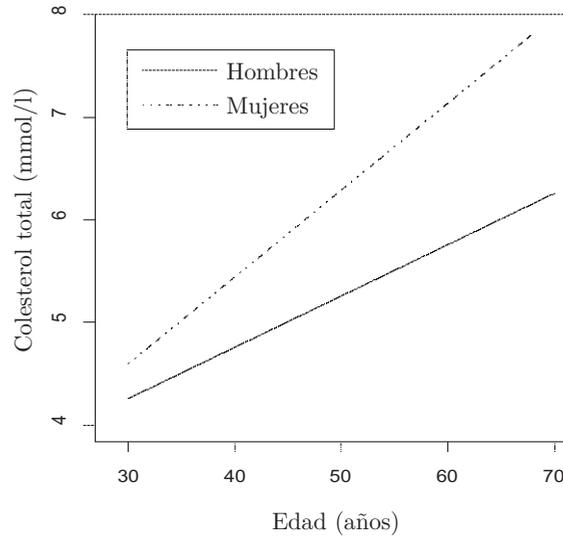


Figura 2.6: Análisis de regresión lineal de la relación entre la edad y el colesterol con interacción entre la edad y el género.

cada médico, como se muestra en la figura 2.7.

En un análisis de regresión estándar esto se hace añadiendo al modelo términos de interacción entre la edad y variables dummy que representan a los distintos médicos, es decir, para m médicos se tiene que

$$\begin{aligned} \text{Colesterol total} = & \beta_0 + \beta_1 \text{ edad} + \beta_2 \text{ dummyMD}_1 + \cdots \\ & + \beta_m \text{ dummyMD}_{m-1} + \beta_{m+1} (\text{dummyMD}_1 \times \text{edad} \\ & + \cdots + \beta_{2m-1} (\text{dummyMD}_{m-1} \times \text{edad}) + \varepsilon, \end{aligned}$$

donde de β_{m+1} hasta β_{2m-1} son los coeficientes de regresión correspondientes a las interacciones entre las variables dummy que representan a cada médico y la edad.

Sin embargo se puede observar que, de nuevo, esto es muy poco eficiente, ya que no hay interés en los términos constantes ni en las pendientes individuales por cada médico. Para esto, el análisis multinivel brinda

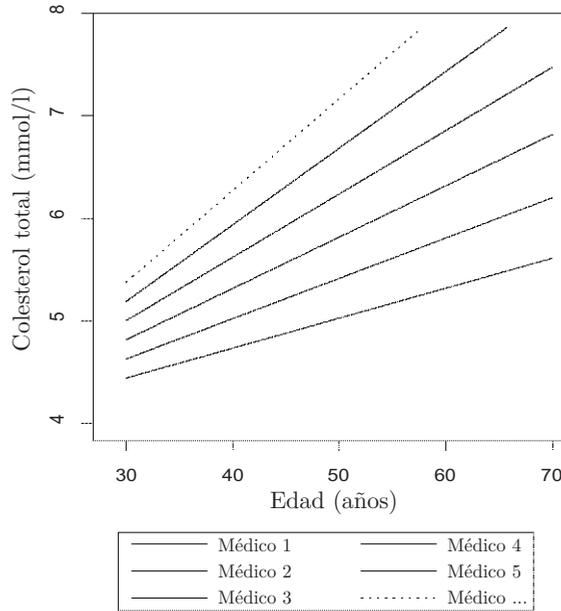
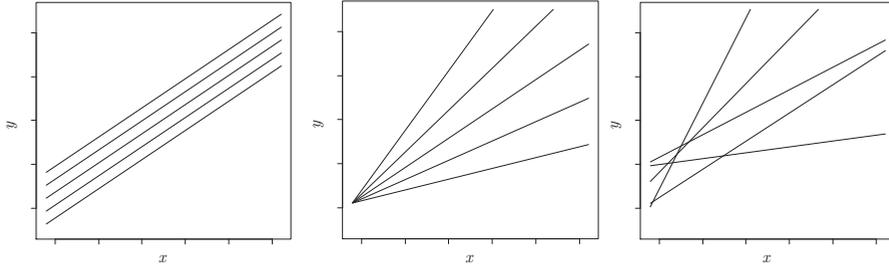


Figura 2.7: *Análisis de regresión lineal de la relación entre la edad y el colesterol total con interacción entre edad y médico.*

una solución muy eficiente. De manera similar a como se planteó para los términos constantes de la regresión, se puede pensar para las pendientes, es decir, estimando la varianza de las pendientes. Es decir, se considera que las pendientes también son aleatorias.

2.3. Coeficientes aleatorios

Cuando se tienen datos que están agrupados en jerarquías que incluyen indicadores por grupos, se denomina un *modelo de intercepto aleatorio*, porque puede ser interpretado como modelo con un término constante de regresión distinto para cada grupo. En la figura 2.8(a) se muestra un modelo con una variable predictora continua x e indicadores



(a) *Interceptos aleatorios.* (b) *Pendientes aleatorias.* (c) *Interceptos y pendientes aleatorios.*

Figura 2.8: *Distintos modelos con coeficientes aleatorios.*

para $J = 5$ grupos. El modelo puede ser escrito como una regresión con 6 variables predictoras o, equivalentemente, una regresión con dos variables predictoras (x y el término constante), con el término constante variable por cada grupo:

$$\text{modelo de interceptos aleatorios: } y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i.$$

Otra opción, que se muestra en la figura 2.8(b), es que la pendiente varíe y que el término constante se mantenga fijo:

$$\text{modelo de pendientes aleatorias: } y_i = \alpha + \beta_{j[i]} x_i + \varepsilon_i.$$

Finalmente, como se muestra en la figura 2.8(c), es posible que tanto el término constante como la pendiente sean aleatorios:

$$\text{modelo de intercepto y pendiente aleatorios: } y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \varepsilon_i.$$

2.3.1. Estructuras de dos niveles

2.3.1.1. Modelo de intercepto aleatorio sin variables predictoras

La regresión multinivel puede verse como un punto intermedio entre los dos extremos de excluir un predictor categórico del modelo (*comple-*

te pooling) y estimar modelos separados para cada nivel del predictor categórico (*no pooling*).

Supóngase que se tiene una muestra que está clasificada en J grupos y que se quiere estimar su distribución. Una estimación puede ser el promedio de todos los datos sin importar las clases, esta estimación ignorará la variación entre las clases. Por otro lado puede calcularse una segunda estimación calculando el promedio de cada clase, sin embargo, en este caso podría estar sobreestimándose la variación de las clases.

Supóngase que se tiene una muestra clasificada en J grupos donde el grupo j tiene tamaño n_j , el elemento i de la clase j , y_{ij} , $j = 1, 2, \dots, J$; $i = 1, 2, \dots, n_j$, entonces, el modelo es:

$$y_{ij} = \beta + u_j + \varepsilon_{ij}, \quad (2.1)$$

con $i_j = 1, 2, \dots, n_j$, donde $\beta \in \mathbb{R}$, $u_j \sim \mathcal{N}(0, \sigma_u^2)$ y $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

Se desea estimar β ya que en este caso este valor caracteriza la distribución de los elementos de la muestra. Para simplificar la notación se escribirá $y_{i_j} = y_{ij}$ y $\varepsilon_{i_j} = \varepsilon_{ij}$ cuando no exista ninguna confusión.

Otra manera de representar este modelo es

$$y_{ij} = \beta_j + \varepsilon_{ij}.$$

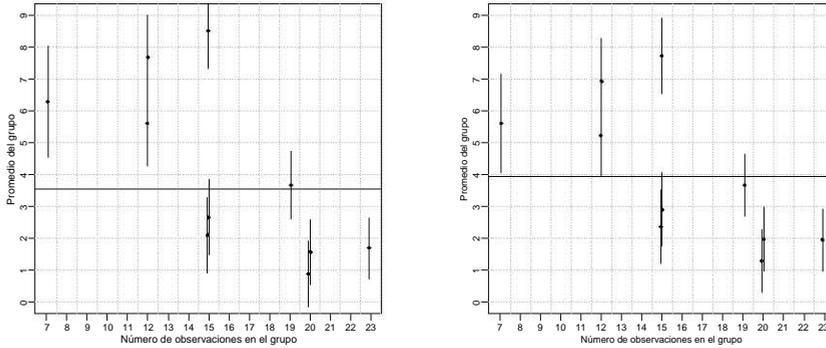
Donde $\beta_j = \beta + u_j$ es el nivel “medio” asociado a la clase j . Además, se puede observar que $\beta_j \sim \mathcal{N}(\beta, \sigma_u^2)$.

2.3.1.1.1. Ejemplo. Se simularon 158 datos contenidos en 10 clases, como se muestra en el cuadro 2.1, de tal modo que $\beta = 4$, $\sigma_u = 2$ y $\sigma_\varepsilon = 5$. En el modelo *complete pooling*, es decir, ignorando la información de las clases y haciendo el promedio de todas las observaciones, obtendremos que $\tilde{\beta}_{pooled} \approx 3.550418$, que podría pensarse que no es una estimación muy adecuada. Por otro lado, en el modelo *no pooling*, es decir, estimado cada clase por separado se obtendrá un estimador por cada una de las clases, de modo que

Y	cls	Y	cls	Y	cls	Y	cls	Y	cls	Y	cls	Y	cls	Y	cls
-3.021	1	1.74	1	3.047	3	5.608	4	14.296	5	8.176	6	1.999	8	-2.113	9
7.61	1	-4.521	1	10.497	3	7.78	4	2.58	6	13.689	7	3.353	8	4.241	9
-0.147	1	-1.376	1	9.03	3	-0.463	4	-1.17	6	4.09	7	11.181	8	-1.245	9
2.464	1	4.964	2	-3.702	3	10.073	4	-3.984	6	8.795	7	-1.758	8	-2.729	10
-0.62	1	11.109	2	6.17	3	11.4	4	6.498	6	-2.418	7	-0.591	8	9.414	10
8.164	1	8.091	2	1.268	3	8.476	4	-1.018	6	6.35	7	5.992	8	4.884	10
5.138	1	2.708	2	-5.807	3	8.534	5	-2.369	6	7.619	7	-0.484	8	0.541	10
8.212	1	7.858	2	3.77	3	11.921	5	-2.96	6	5.923	7	-0.361	8	1.774	10
7.297	1	-0.071	2	-3.587	3	5.611	5	3.105	6	-2.956	8	-2.716	9	4.513	10
0.914	1	5.168	2	-0.518	3	12.231	5	12.982	6	3.327	8	-1.075	9	3.72	10
-7.546	1	4.414	2	3.335	3	11.523	5	-0.306	6	-7.623	8	10.46	9	-0.577	10
5.973	1	-0.779	2	4.735	3	0.289	5	-5.319	6	1.784	8	5.167	9	-0.562	10
1.599	1	9.723	2	-1.701	3	6.758	5	-4.555	6	4.489	8	7.367	9	-2.693	10
-3.007	1	6.137	2	0.598	3	-0.332	5	6.926	6	-2.467	8	13.494	9	3.014	10
5.236	1	8.02	2	7.848	4	5.499	5	4.55	6	-3.082	8	-4.668	9	2.57	10
4.034	1	5.452	3	3.468	4	13.054	5	-2.292	6	4.473	8	6.216	9	2.767	10
-2.604	1	15.209	3	12.849	4	6.708	5	6.213	6	-0.105	8	1.861	9	4.206	10
-2.186	1	9.027	3	12.318	4	11.556	5	2.03	6	9.485	8	3.177	9	0.755	10
-2.748	1	1.806	3	8.057	4	8.923	5	1.026	6	3.85	8	-1.298	9		
8.542	1	11.152	3	4.877	4	11.057	5	-12.434	6	1.014	8	1.065	9		

Cuadro 2.1: Muestra simulada para el modelo de intercepto aleatorio sin variables regresoras.

$$\begin{aligned} \widehat{\beta}_{1_{\text{no pooled}}} &\approx 1.70204 \\ \widehat{\beta}_{2_{\text{no pooled}}} &\approx 5.61183 \\ \widehat{\beta}_{3_{\text{no pooled}}} &\approx 3.67268 \\ \widehat{\beta}_{4_{\text{no pooled}}} &\approx 7.69092 \\ \widehat{\beta}_{5_{\text{no pooled}}} &\approx 8.50853 \\ \widehat{\beta}_{6_{\text{no pooled}}} &\approx 0.88395 \\ \widehat{\beta}_{7_{\text{no pooled}}} &\approx 6.29257 \\ \widehat{\beta}_{8_{\text{no pooled}}} &\approx 1.576 \\ \widehat{\beta}_{9_{\text{no pooled}}} &\approx 2.6622 \\ \widehat{\beta}_{10_{\text{no pooled}}} &\approx 2.10647 \end{aligned}$$



(a) Modelo Pooling y modelo No-Pooling (b) Modelo Multinivel

Figura 2.9: En la figura (a) los puntos muestran los promedios de cada clase sin tomar en cuenta la información de las demás clases y se muestra con una línea vertical su desviación estándar. La línea vertical representa el promedio de todos los datos de la muestra ignorando la información de su separación por grupos. En la figura (b) los puntos muestran el ajuste por grupos que se hace en el análisis multinivel con sus desviaciones estándar y la línea vertical el estimador de la media de toda la población tomando en cuenta toda la información.

lo que tampoco parece dar un buen estimador de β . Por otro lado, el estimador utilizando análisis multinivel, será $\hat{\beta}_{multinivel} \approx 3.955913$ con $\hat{\sigma}_u \approx 2.4562$ y $\hat{\sigma}_\epsilon \approx 4.7253$. La comparativa visual se muestra en la figura 2.9.

2.3.1.2. Intercepto aleatorio con una variable predictora

El siguiente paso en el modelo, es la introducción de una variable predictora x , es decir, un modelo lineal en el cual el intercepto es considerado aleatorio y se considera que tiene una distribución normal. Supóngase que se tiene una muestra clasificada en J grupos donde el gru-

X	Y	cls	X	Y	cls	X	Y	cls	X	Y	cls	X	Y	cls
5.203	-3.738	1	0.436	-1.815	2	6.263	10.745	4	4.573	37.159	6	7.067	38.909	8
14.957	30.898	1	14.955	57.747	2	8.606	18.784	4	1.653	26.666	6	6.187	36.861	9
7.634	2.342	1	9.302	35.654	2	6.648	29.549	5	3.572	51.927	7	14.158	64.347	9
3.444	-14.51	1	1.607	10.502	3	11.928	48.937	5	10.556	77.995	7	12.031	61.053	9
9.444	9.572	1	14.538	60.762	3	4.635	19.07	5	9.84	70.78	7	5.917	32.809	10
14.822	33.423	1	10.012	40.704	3	4.135	21.354	5	3.521	46.433	7	2.553	18.537	10
13.793	55.425	2	12.979	50.065	3	8.352	33.878	5	5.857	57.924	7	8.31	40.293	10
7.662	29.189	2	4.066	14.953	3	6.89	45.605	6	2.813	43.42	7	2.982	21.861	10
12.045	49.257	2	3.735	2.629	4	2.802	28.626	6	2.533	44.741	7	14.936	67.717	10
9.918	38.487	2	6.501	10.262	4	11.082	66.409	6	2.409	39.56	7	11.087	53.232	10
5.3	17.957	2	4.548	4.279	4	13.097	76.671	6	9.707	74.007	7	7.637	36.259	10
2.034	5.19	2	7.096	14.954	4	3.89	36.613	6	4.684	52.068	7	8.413	41.688	10
2.363	6.582	2	0.179	-13.414	4	14.358	76.559	6	14.88	70.587	8	6.918	34.682	10

Cuadro 2.2: Muestra simulada para el modelo de intercepto aleatorio con una variable regresora.

po j tiene tamaño n_j . En este caso, el elemento i de la clase j , y_i , $j = 1, 2, \dots, J; i = 1, 2, \dots, n_j$, será de la siguiente forma:

$$y_i = \beta_{0j} + \beta_1 x + \varepsilon_i,$$

donde $\beta_{0j} \sim \mathcal{N}(\beta_0, \sigma_u^2)$ y $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Otra manera de expresar este modelo es

$$y_i = \beta_0 + u_j + \beta_1 x + \varepsilon_i,$$

donde $u_j \sim \mathcal{N}(0, \sigma_u^2)$ y $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ son independientes.

2.3.1.2.1. **Ejemplo.** Se simularon 65 datos contenidos en 10 clases, como se muestra en el cuadro 2.2.

En la simulación, $\beta_0 = 4$, $\beta_1 = 4$, $\sigma_u = 20$ y $\sigma_\varepsilon = 2$. Haciendo la estimación ignorando la información de la separación por clases, se obtiene que $\widehat{\beta}_0 \text{ pooled} \approx 8.790640$ y $\widehat{\beta}_1 \text{ pooled} \approx 3.559857$. En el modelo *no-pooling* se obtiene que

$\widehat{\beta}_0$ 1 _{no pooled} ≈ -26.905531	$\widehat{\beta}_1$ 1 _{no pooled} ≈ 3.953232
$\widehat{\beta}_0$ 2 _{no pooled} ≈ -3.515473	$\widehat{\beta}_1$ 2 _{no pooled} ≈ 4.226143
$\widehat{\beta}_0$ 3 _{no pooled} ≈ 1.823689	$\widehat{\beta}_1$ 3 _{no pooled} ≈ 3.885643
$\widehat{\beta}_0$ 4 _{no pooled} ≈ -13.188425	$\widehat{\beta}_1$ 4 _{no pooled} ≈ 3.806271
$\widehat{\beta}_0$ 5 _{no pooled} ≈ 3.931460	$\widehat{\beta}_1$ 5 _{no pooled} ≈ 3.729360
$\widehat{\beta}_0$ 6 _{no pooled} ≈ 18.470271	$\widehat{\beta}_1$ 6 _{no pooled} ≈ 4.225655
$\widehat{\beta}_0$ 7 _{no pooled} ≈ 32.790093	$\widehat{\beta}_1$ 7 _{no pooled} ≈ 4.161934
$\widehat{\beta}_0$ 8 _{no pooled} ≈ 10.255675	$\widehat{\beta}_1$ 8 _{no pooled} ≈ 4.054525
$\widehat{\beta}_0$ 9 _{no pooled} ≈ 15.287811	$\widehat{\beta}_1$ 9 _{no pooled} ≈ 3.595181
$\widehat{\beta}_0$ 10 _{no pooled} ≈ 8.636717	$\widehat{\beta}_1$ 10 _{no pooled} ≈ 3.917612

El estimador utilizando análisis multinivel, será $\widehat{\beta}_0$ *multinivel* ≈ 3.951198 , $\widehat{\beta}_1$ *multinivel* ≈ 4.044697 con $\widehat{\sigma}_u \approx 17.0368$ y $\widehat{\sigma}_\varepsilon \approx 1.8952$. La comparativa visual se muestra en la figura 2.10.

2.3.1.3. Intercepto y pendiente aleatorios con una variable predictora

Ahora se considerará el caso en el que los interceptos y la pendiente pueden variar por grupo. Supóngase que se tiene una muestra clasificada en J grupos donde el grupo j tiene tamaño n_j , y_{ij} es el elemento i de la clase j , $j = 1, 2, \dots, J$; $i = 1, 2, \dots, n_j$, entonces, el modelo es:

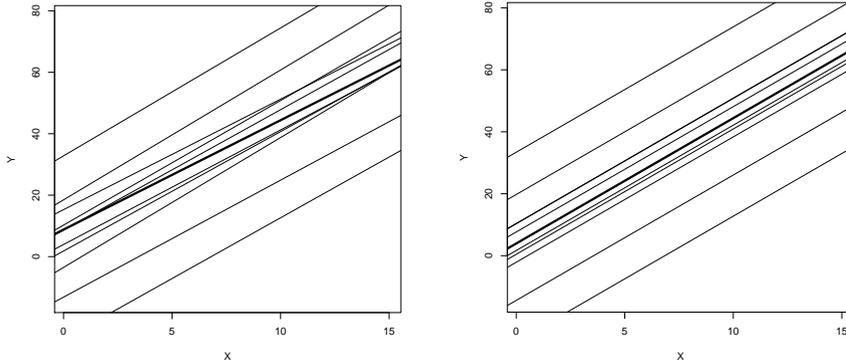
$$y_{ij} = \beta_{0j} + \beta_{1j}x + \varepsilon_{ij}. \quad (2.2)$$

Donde

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \right).$$

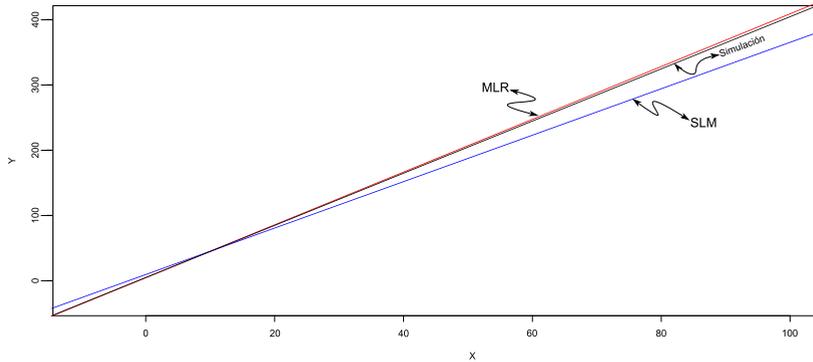
Al no ser necesariamente independientes β_{0j} y β_{1j} , existe un coeficiente de correlación entre estos parámetros que es $\rho_{u_0 u_1}$. Otra forma de expresar este modelo, teniendo de manera expresa la parte aleatoria de y_{ij} , es

$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x + \varepsilon_{ij} = \beta_0 + \beta_1 x + (u_{0j} + u_{1j}x + \varepsilon_{ij}).$$



(a) Modelo Pooling y modelo No-Pooling

(b) Modelo Multinivel



(c) Modelos Comparados

Figura 2.10: En la figura (a) la línea más gruesa es el modelo complete pooling y las demás líneas son los modelos estimados de cada una de las 10 clases. En la figura (b) la línea más gruesa es la estimación del modelo multinivel y las demás líneas son las estimaciones de las clases ajustadas a la información de toda la población. En la figura (c) se muestra una comparación de la estimación lineal usual (SLM), la estimación con el modelo multinivel (MLR) y la línea base para la simulación (Simulación).

X	Y	cls	X	Y	cls	X	Y	cls	X	Y	cls	X	Y	cls
9.936	67.793	1	0.57	-9.969	3	11.585	-8.304	4	1.936	2.723	6	3.054	42.994	8
3.221	15.572	1	4.174	-43.562	3	6.319	7.175	4	5.909	32.737	6	5.917	75.603	8
10.766	68.421	1	3.813	-43.405	3	1.741	19.354	4	4.586	25.904	6	4.043	16.312	9
4.794	27.535	1	10.465	-104.312	3	2.848	13.091	5	3.05	18.434	7	5.197	18.175	9
7.193	46.82	1	13.379	-132.749	3	11.475	29.933	5	11.24	40.998	7	7.102	20.176	9
5.435	32.161	1	12.579	-123.348	3	2.544	12.088	5	0.628	6.842	7	8.94	21.085	9
8.4	55.954	1	11.918	-6.641	4	11.488	29.467	5	13.93	45.601	7	13.968	29.396	9
9.187	98.34	2	10.993	-8.333	4	14.089	90.098	6	8.258	30.263	7	9.908	19.959	9
4.771	59.013	2	11.666	-6.954	4	14.105	92.987	6	12.453	42.525	7	0.127	9.388	9
11.246	118.506	2	11.26	-5.521	4	14.549	95.159	6	13.552	46.82	7	10.239	100.003	10
1.553	23.922	2	11.802	-6.188	4	4.865	22.741	6	8.636	37.018	7	2.531	19.471	10
8.811	95.029	2	14.608	-15.578	4	13.605	89.512	6	9.479	115.737	8	9.03	88.617	10
12.016	-119.87	3	8.683	-0.299	4	11.196	71.411	6	2.917	42.855	8	10.465	100.192	10

Cuadro 2.3: Muestra simulada para el modelo con intercepto y pendiente aleatorios.

Donde

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \right) \quad (2.3)$$

y

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

son independientes.

2.3.1.3.1. Ejemplo. Se simularon 65 datos contenidos en 10 clases, como se muestra en el cuadro 2.3, de tal modo que $\beta_0 = 4$, $\beta_1 = 4$, independientes y $\sigma_{u_0} = 10$, $\sigma_{u_1} = 5$, $\sigma_\varepsilon = 2$. Haciendo la estimación ignorando la información de la separación por clases, se obtiene que $\hat{\beta}_0 \text{ pooled} \approx 17.939408$ y $\hat{\beta}_1 \text{ pooled} \approx 1.000137$. En el modelo *no-pooling* se obtiene que

$$\begin{array}{ll}
\widehat{\beta}_0 \text{ 1}_{\text{no pooled}} \approx -7.382533 & \widehat{\beta}_1 \text{ 1}_{\text{no pooled}} \approx 7.356191 \\
\widehat{\beta}_0 \text{ 2}_{\text{no pooled}} \approx 10.522905 & \widehat{\beta}_1 \text{ 2}_{\text{no pooled}} \approx 9.620880 \\
\widehat{\beta}_0 \text{ 3}_{\text{no pooled}} \approx -5.154519 & \widehat{\beta}_1 \text{ 3}_{\text{no pooled}} \approx -9.494234 \\
\widehat{\beta}_0 \text{ 4}_{\text{no pooled}} \approx 23.782920 & \widehat{\beta}_1 \text{ 4}_{\text{no pooled}} \approx -2.675796 \\
\widehat{\beta}_0 \text{ 5}_{\text{no pooled}} \approx 7.333373 & \widehat{\beta}_1 \text{ 5}_{\text{no pooled}} \approx 1.948352 \\
\widehat{\beta}_0 \text{ 6}_{\text{no pooled}} \approx -10.442093 & \widehat{\beta}_1 \text{ 6}_{\text{no pooled}} \approx 7.275470 \\
\widehat{\beta}_0 \text{ 7}_{\text{no pooled}} \approx 7.841288 & \widehat{\beta}_1 \text{ 7}_{\text{no pooled}} \approx 2.868004 \\
\widehat{\beta}_0 \text{ 8}_{\text{no pooled}} \approx 9.438124 & \widehat{\beta}_1 \text{ 8}_{\text{no pooled}} \approx 11.205902 \\
\widehat{\beta}_0 \text{ 9}_{\text{no pooled}} \approx 10.128989 & \widehat{\beta}_1 \text{ 9}_{\text{no pooled}} \approx 1.290212 \\
\widehat{\beta}_0 \text{ 10}_{\text{no pooled}} \approx -6.434780 & \widehat{\beta}_1 \text{ 10}_{\text{no pooled}} \approx 10.352460
\end{array}$$

El estimador utilizando análisis multinivel, será $\widehat{\beta}_0 \text{ multinivel} \approx 3.979465$, $\widehat{\beta}_1 \text{ multinivel} \approx 3.972333$ con $\widehat{\sigma}_{u_0} \approx 10.7101$, $\widehat{\sigma}_{u_1} \approx 6.5384$, $\widehat{\rho}_{u_0 u_1} \approx -0.171$ y $\widehat{\sigma}_\varepsilon \approx 1.7310$. La comparativa se muestra en la figura 2.11.

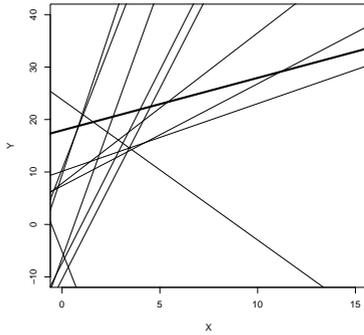
2.3.1.4. Intercepto y pendientes aleatorios con h variables predictoras

En el caso en el que los interceptos y la pendiente pueden variar por grupo se hace de manera similar que el caso anterior, es decir, supóngase que se tiene una muestra clasificada en J grupos donde el grupo j tiene tamaño n_j , el elemento i de la clase j , y_{ij} , $j = 1, 2, \dots, J$; $i = 1, 2, \dots, n_j$, donde el valor de y_{ij} está afectado por h variables de predicción, x_1, x_2, \dots, x_h , entonces, el modelo es:

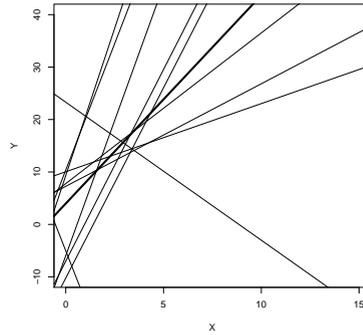
$$y_{ij} = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \dots + \beta_{hj}x_h + \varepsilon_{ij}. \quad (2.4)$$

Donde

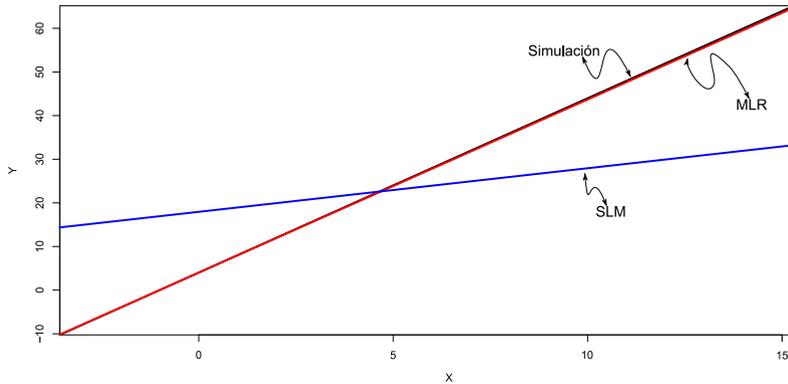
$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{hj} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{hj} \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \cdots & \rho_{u_0 u_h} \sigma_{u_0} \sigma_{u_h} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 & \cdots & \rho_{u_1 u_h} \sigma_{u_1} \sigma_{u_h} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{u_0 u_h} \sigma_{u_0} \sigma_{u_h} & \rho_{u_1 u_h} \sigma_{u_1} \sigma_{u_h} & \cdots & \sigma_{u_h}^2 \end{pmatrix} \right).$$



(a) *Modelo Pooling y modelo No-Pooling*



(b) *Modelo Multinivel*



(c) *Modelos Comparados*

Figura 2.11: *En la figura (a) La línea más gruesa es el modelo complete pooling y las demás líneas son los modelos estimados de cada una de las 10 clases. En la figura (b) la línea más gruesa es la estimación del modelo multinivel y las demás líneas son las estimaciones de las clases ajustadas a la información de toda la población. En la figura (c) se muestra una comparación de la estimación lineal usual (SLM), la estimación con el modelo multinivel (MLR) y la línea base para la simulación (Simulación).*

Al no ser necesariamente independientes los parámetros β_{kj} ($k = 1, 2, \dots, h$), existe un coeficiente de correlación entre estos parámetros que es $\rho_{u_k u_l}$ con $l = 1, 2, \dots, h$ y $k \neq l$. Otra forma de expresar este modelo, teniendo de manera expresa la parte aleatoria de y_{ij} , es

$$y_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_h x_h + (u_{0j} + u_{1j} x_1 + u_{2j} x_2 + \dots + u_{hj} x_h + \varepsilon_{ij}).$$

Donde

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{hj} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \cdots & \rho_{u_0 u_h} \sigma_{u_0} \sigma_{u_h} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 & \cdots & \rho_{u_1 u_h} \sigma_{u_1} \sigma_{u_h} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{u_0 u_h} \sigma_{u_0} \sigma_{u_h} & \rho_{u_1 u_h} \sigma_{u_1} \sigma_{u_h} & \cdots & \sigma_{u_h}^2 \end{pmatrix} \right) \quad (2.5)$$

y

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

son independientes.

2.3.1.5. Intercepto y pendiente aleatorios con una variable predictora en el primer nivel y una variable predictora al nivel de grupo

El modelo (2.2) se puede extender a través de la introducción de una variable predictora en el nivel superior z , de esta manera, el modelo constará de la ecuación de regresión del primer nivel que será

$$y_{ij} = \beta_{0j} + \beta_{1j} x + \varepsilon_{ij}$$

y las ecuaciones del segundo nivel que serán

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} z_j + u_{1j}. \end{aligned}$$

Donde la distribución de las variables u_{0j} y u_{1j} será la misma que en (2.3), es decir,

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \right).$$

2.3.1.6. Intercepto y pendiente aleatorios con h variables predictoras en el primer nivel y una variable predictora al nivel de grupo

Es una extensión del modelo (2.4), añadiendo la variable predictora z en el segundo nivel, entonces la ecuación de primer nivel será la misma que en (2.4)

$$y_{ij} = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \cdots + \beta_{hj}x_h + \varepsilon_{ij}$$

y las ecuaciones del segundo nivel que serán

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}z_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}z_j + u_{1j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}z_j + u_{2j} \\ &\vdots \\ \beta_{hj} &= \gamma_{h0} + \gamma_{h1}z_j + u_{hj}. \end{aligned}$$

Donde la distribución de las variables u_{0j} y u_{1j} será la misma que en (2.5), es decir,

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{hj} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0u_1}\sigma_{u_0}\sigma_{u_1} & \cdots & \rho_{u_0u_h}\sigma_{u_0}\sigma_{u_h} \\ \rho_{u_0u_1}\sigma_{u_0}\sigma_{u_1} & \sigma_{u_1}^2 & \cdots & \rho_{u_1u_h}\sigma_{u_1}\sigma_{u_h} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{u_0u_h}\sigma_{u_0}\sigma_{u_h} & \rho_{u_1u_h}\sigma_{u_1}\sigma_{u_h} & \cdots & \sigma_{u_h}^2 \end{pmatrix} \right).$$

2.3.1.7. Intercepto y pendiente aleatorios con h variables predictoras en el primer nivel y L variables predictoras al nivel de grupo

Es una extensión del modelo anterior, por lo tanto, la ecuación del primer nivel será la misma que en (2.4), es decir,

$$y_{ij} = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \cdots + \beta_{hj}x_h + \varepsilon_{ij}.$$

Ahora, defínase

$$\bar{\gamma}_l = (\gamma_{l0}, \gamma_{l1}, \gamma_{l2}, \cdots, \gamma_{lh}),$$

con $l = 0, 1, 2, \dots, L$. También se define

$$Z = \begin{pmatrix} 1 \\ z_1 \\ z_2 \\ \vdots \\ z_h \end{pmatrix}.$$

Entonces, las ecuaciones del segundo nivel serán

$$\begin{aligned} \beta_{0j} &= \bar{\gamma}_0 Z + u_{0j} \\ \beta_{1j} &= \bar{\gamma}_1 Z + u_{1j} \\ \beta_{2j} &= \bar{\gamma}_2 Z + u_{2j} \\ &\vdots \\ \beta_{hj} &= \bar{\gamma}_h Z + u_{hj}. \end{aligned}$$

Y la distribución de las variables u_{kj} será como en (2.5).

2.3.2. Estructuras con más de dos niveles

2.3.2.1. Cuatro niveles sin variables predictoras

$$\begin{aligned} y_{ijkl} &= \beta_{0,jkl} + \varepsilon_{ijkl} \\ \beta_{0,jkl} &= \beta_{0,kl} + \varepsilon_{0,jkl} \\ \beta_{0,kl} &= \beta_{0,l} + \varepsilon_{0,kl} \\ \beta_{0,l} &= \beta_0 + \varepsilon_{0,l} \\ \varepsilon_{ijkl} &\sim \mathcal{N}(0, \sigma_1^2) \\ \varepsilon_{0,jkl} &\sim \mathcal{N}(0, \sigma_2^2) \\ \varepsilon_{0,kl} &\sim \mathcal{N}(0, \sigma_3^2) \\ \varepsilon_{0,l} &\sim \mathcal{N}(0, \sigma_4^2) \end{aligned}$$

2.3.2.2. Tres niveles con una variable predictora en cada nivel

Las ecuaciones en cada uno de los niveles serán:

$y_{ijk} = \beta_{0jk} + \beta_{1jk}x_{ijk} + \varepsilon_{ijk}$	1 ^{er} Nivel
$\beta_{0jk} = \beta_{00k} + \beta_{001}x_k + \varepsilon_{0jk}$	2 ^{do} Nivel
$\beta_{1jk} = \beta_{10k} + \beta_{11k}x_k + \varepsilon_{1jk}$	3 ^{er} Nivel
$\beta_{00k} = \beta_{000} + \beta_{001}x_k + \varepsilon_{00k}$	
$\beta_{01k} = \beta_{010} + \beta_{011}x_k + \varepsilon_{01k}$	
$\beta_{10k} = \beta_{100} + \beta_{101}x_k + \varepsilon_{10k}$	
$\beta_{11k} = \beta_{110} + \beta_{111}x_k + \varepsilon_{11k}$	

donde

$$\begin{aligned} \varepsilon_{ijk} &\sim \mathcal{N}(0, \sigma^2) \\ \begin{pmatrix} \varepsilon_{0jk} \\ \varepsilon_{1jk} \end{pmatrix} &\sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_{0,1}\sigma_0\sigma_1 \\ 0 & \sigma_1^2 \end{pmatrix} \right) \\ \begin{pmatrix} \varepsilon_{00k} \\ \varepsilon_{01k} \\ \varepsilon_{10k} \\ \varepsilon_{11k} \end{pmatrix} &\sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{00}^2 & \rho_{00,01}\sigma_{00}\sigma_{01} & \rho_{00,10}\sigma_{00}\sigma_{10} & \rho_{00,11}\sigma_{00}\sigma_{11} \\ \rho_{00,01}\sigma_{00}\sigma_{01} & \sigma_{01}^2 & \rho_{01,10}\sigma_{01}\sigma_{10} & \rho_{01,11}\sigma_{01}\sigma_{11} \\ \rho_{00,10}\sigma_{00}\sigma_{10} & \rho_{01,10}\sigma_{01}\sigma_{10} & \sigma_{10}^2 & \rho_{10,11}\sigma_{10}\sigma_{11} \\ \rho_{00,11}\sigma_{00}\sigma_{11} & \rho_{01,11}\sigma_{01}\sigma_{11} & \rho_{10,11}\sigma_{10}\sigma_{11} & \sigma_{11}^2 \end{pmatrix} \right) \end{aligned}$$

Entonces, la ecuación quedará de la siguiente manera

$$\underbrace{\left[\underbrace{\left(\underbrace{\beta_{00k}}_{\beta_{00}(x_k)} + \underbrace{\beta_{001}x_k + \varepsilon_{00k}}_{\beta_{01}(x_k)} \right)}_{\beta_{0j}(x_k)} \right]}_{\beta_{0jk}} + \underbrace{\left[\underbrace{\left(\underbrace{\beta_{010} + \beta_{011}x_k + \varepsilon_{01k}}_{\beta_{10}(x_k)} + \underbrace{x_{jk} + \varepsilon_{0j}}_{\beta_{11}(x_k)} \right)}_{\beta_{1j}(x_k)} \right]}_{\beta_{1jk}} + \underbrace{\left[\underbrace{\left(\underbrace{\beta_{100} + \beta_{101}x_k + \varepsilon_{10k}}_{\beta_{10}(x_k)} + \underbrace{\left(\beta_{110} + \beta_{111}x_k + \varepsilon_{11k} \right)}_{\beta_{11}(x_k)} \right)}_{\beta_{1j}(x_k)} + \underbrace{x_{jk} + \varepsilon_{1j}}_{\beta_{1j}(x_k)} \right]}_{\beta_{1jk}} + \underbrace{\varepsilon_{ijk}}_{\beta_{ijk}(x_k, x_k)}$$

2.4. Estimación de parámetros

2.4.1. El modelo de componentes de la varianza

En el modelo $\beta_{0j} + \beta_{1j} + \varepsilon_{ij}$ se requiere la estimación de dos componentes fijos β_0 y β_1 y otros cuatro parámetros σ_ε^2 , $\sigma_{u_0}^2$, $\sigma_{u_1}^2$ y $\sigma_{u_{01}} = \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1}$. Sin embargo, considérese primero el modelo más sencillo de dos niveles, que sólo incluye los componentes fijos y los parámetros σ_ε^2 y $\sigma_{u_0}^2$, es decir, el modelo (2.1), que puede ser expresado como

$$y_{ij} = \beta + u_j + \varepsilon_{ij} \quad (2.6)$$

donde

$$\text{Var}(\varepsilon_{ij}) = \sigma_\varepsilon^2, \quad \text{Var}(u_j) = \sigma_{u_0}^2.$$

Se denomina modelo de componentes de la varianza ya que la varianza de la respuesta es

$$\text{Var}(y_{ij}) = \text{Var}(\beta + u_j + \varepsilon_{ij}) = \text{Var}(u_j + \varepsilon_{ij}) = \sigma_{u_0}^2 + \sigma_\varepsilon^2.$$

Es decir, la varianza de la respuesta es la suma de la varianza del primer y del segundo nivel, donde los errores de ambos niveles se suponen independientes entre sí. Ahora si se toman dos elementos distintos pertenecientes a la misma clase, por ejemplo el elemento i_1 y el elemento i_2 , la covarianza entre ambos estará dada por

$$\begin{aligned} \text{cov}(y_{i_1 j}, y_{i_2 j}) &= \text{cov}(\beta + u_j + \varepsilon_{i_1 j}, \beta + u_j + \varepsilon_{i_2 j}) = \text{cov}(u_j, u_j) \\ &= \text{Var}(u_j) = \sigma_{u_0}^2. \end{aligned}$$

Como los errores del primer nivel son independientes, entonces la correlación entre los elementos de una misma clase será

$$\rho = \frac{\text{cov}(y_{i_1 j}, y_{i_2 j})}{\sqrt{\text{Var}(y_{i_1 j}) \text{Var}(y_{i_2 j})}} = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_\varepsilon^2}.$$

Esta expresión es conocida como el *coeficiente de correlación intraclass* o como el *coeficiente de correlación inter nivel 2*.

De lo anterior, se tiene que la matriz de covarianzas para tres elementos de un mismo grupo será

$$\begin{pmatrix} \sigma_{u_0}^2 + \sigma_\varepsilon^2 & \sigma_{u_0}^2 & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \sigma_{u_0}^2 + \sigma_\varepsilon^2 & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \sigma_{u_0}^2 & \sigma_{u_0}^2 + \sigma_\varepsilon^2 \end{pmatrix}.$$

Para dos clases con 2 y tres elementos, la matriz de covarianzas será de la forma

$$V = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \quad (2.7)$$

donde

$$A = \begin{pmatrix} \sigma_{u_0}^2 + \sigma_\varepsilon^2 & \sigma_{u_0}^2 & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \sigma_{u_0}^2 + \sigma_\varepsilon^2 & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \sigma_{u_0}^2 & \sigma_{u_0}^2 + \sigma_\varepsilon^2 \end{pmatrix},$$

$$B = \begin{pmatrix} \sigma_{u_0}^2 + \sigma_\varepsilon^2 & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \sigma_{u_0}^2 + \sigma_\varepsilon^2 \end{pmatrix}.$$

La estructura de bloques en la diagonal de la matriz refleja el hecho de que la covarianza entre elementos de distintas clase es cero y se extiende fácilmente a cualquier número de clases y cualquier número de elementos por clase.

Sea $I_{(n)}$ la matriz identidad de $(n \times n)$ y sea $J_{(n)}$ la matriz que contiene 1 en cada una de sus entradas de $(n \times n)$, entonces, se puede expresar la matriz V de la siguiente manera:

$$V = \begin{pmatrix} \sigma_{u_0}^2 J_{(3)} + \sigma_\varepsilon^2 I_{(3)} & 0 \\ 0 & \sigma_{u_0}^2 J_{(2)} + \sigma_\varepsilon^2 I_{(2)} \end{pmatrix}.$$

2.4.2. El modelo general de 2 niveles con coeficientes aleatorios

Se puede extender el modelo (2.6) para incluir más variables explicativas de la siguiente manera

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sum_{h=2}^p \beta_h x_{hij} + (u_{0j} + u_{1j} x_{1ij} + \varepsilon_{ij}).$$

O, equivalentemente,

$$y_{ij} = X_{ij} \beta + \sum_{h=0}^1 u_{hj} z_{hij} + \varepsilon_{ij} z_{0ij},$$

donde se utiliza una nueva formulación de las variables explicativas en la parte aleatoria del modelo y se escribe de manera más general como

$$\begin{aligned} Z &= \{Z_0 Z_1\} \\ Z_0 &= \{1\} \quad \text{un vector de 1's} \\ Z_1 &= \{x_{1ij}\}. \end{aligned}$$

Las variables explicativas para la parte aleatoria del modelo son comúnmente un subconjunto de aquellas en la parte determinista, como en este caso, sin embargo, este no siempre es el caso. También, las variables explicativas pueden ser medidas en cualquiera de los niveles.

En este modelo, con el coeficiente de X_1 aleatorio al segundo nivel, genera la siguiente estructura de bloque para una estructura de dos niveles con dos elementos en el primer nivel. La matriz Ω_2 es la matriz de covarianzas de la pendiente y el intercepto aleatorio. La matriz Ω_1 es la matriz de covarianzas de los elementos del nivel uno, en este nivel, esta matriz será simplemente la varianza del nivel 1.

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

donde

$$\begin{aligned} A &= (\sigma_{u_0}^2 + 2\sigma_{u_{01}}x_{1j} + \sigma_{u_1}^2x_{1j}^2 + \sigma_\varepsilon^2) \\ B &= (\sigma_{u_0}^2 + \sigma_{u_{01}}(x_{1j} + x_{2j}) + \sigma_{u_1}^2x_{1j}x_{2j}) \\ C &= (\sigma_{u_0}^2 + 2\sigma_{u_{01}}x_{2j} + \sigma_{u_1}^2x_{2j}^2 + \sigma_\varepsilon^2) \end{aligned}$$

por lo tanto,

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} = X_j \Omega_2 X_j^T + \begin{pmatrix} \Omega_1 & 0 \\ 0 & \Omega_1 \end{pmatrix}$$

$$X_j = \begin{pmatrix} 1 & x_{1j} \\ 1 & x_{2j} \end{pmatrix}, \quad \Omega_2 = \begin{pmatrix} \sigma_{u_0}^2 & \sigma_{u_{01}} \\ \sigma_{u_{01}} & \sigma_{u_1}^2 \end{pmatrix}, \quad \Omega_1 = \sigma_\varepsilon^2.$$

2.4.3. Estimación por máxima verosimilitud utilizando mínimos cuadrados iterativos generalizados

Considérese el modelo de componentes de la varianza para dos niveles,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + \varepsilon_{ij},$$

donde

$$\text{Var}(\varepsilon_{ij}) = \sigma_\varepsilon^2, \quad \text{Var}(u_{0j}^2) = \sigma_{u_0}^2.$$

Supóngase que se conocen los valores de las varianzas y que por lo tanto se puede construir de manera inmediata la matriz V . Se podría, entonces, aplicar la estimación usual por mínimos cuadrados para obtener la estimación de los coeficientes fijos, esto es

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (2.8)$$

donde, en este caso,

$$X = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n_m m} \end{pmatrix}, \quad Y = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{n_m m} \end{pmatrix}.$$

Con m clases y n_j elementos en la j -ésima clase. Como se supone que los errores son normalmente distribuidos, el estimador de β obtenido en (2.8) es además máximo verosímil.

El procedimiento de estimación es iterativo. Se comenzará utilizando estimadores ‘razonables’ de los parámetros fijos. Estos pueden ser obtenidos por medio de un ajuste inicial por mínimos cuadrados ordinarios (es decir, suponiendo que $\sigma_{u_0}^2 = 0$), de los cuales se obtienen los estimadores β^{OLS} . A partir de esta estimación inicial se obtienen los residuales ‘brutos’

$$\tilde{y}_{ij} = y_{ij} - \hat{\beta}_0^{\text{OLS}} - \hat{\beta}_1^{\text{OLS}} x_{ij}$$

que puede ser representado como

$$\tilde{Y} = \{\tilde{y}_{ij}\}$$

Si los estimadores OLS^1 son insesgados (son de hecho consistentes), el valor esperado de $\tilde{Y}\tilde{Y}^T$ es simplemente V . Se puede reacomodar $\tilde{Y}\tilde{Y}^T$ como un vector apilando las columnas una sobre otra lo cual se escribirá como $\text{vec}(\tilde{Y}\tilde{Y}^T)$ y de manera similar se puede construir el vector $\text{vec}(V)$. Para la estructura que se muestra en la ecuación (2.7) se tienen $3^2 + 2^2 = 13$ elementos. La relación entre estos vectores se puede expresar como el siguiente modelo lineal

$$\begin{pmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{21}\tilde{y}_{11} \\ \vdots \\ \tilde{y}_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{u_0}^2 + \sigma_\varepsilon^2 \\ \sigma_{u_0}^2 \\ \vdots \\ \sigma_{u_0}^2 + \sigma_\varepsilon^2 \end{pmatrix} + R = \sigma_{u_0}^2 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \sigma_\varepsilon^2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} + R, \quad (2.9)$$

donde R es el vector de errores. El lado izquierdo de la ecuación (2.9) es el vector de respuestas en el modelo lineal y el lado derecho contiene dos variables explicativas con coeficientes $\sigma_{u_0}^2$, σ_ε^2 que también deben ser

¹De *Ordinary least squares*

estimados. La estimación involucra la aplicación de mínimos cuadrados generalizados o GLS² utilizando la matriz de covarianzas estimada de $\text{vec}(\tilde{Y}\tilde{Y}^T)$, bajo la hipótesis de normalidad, es decir, $2(V^{-1} \otimes V^{-1})$ donde \otimes es el producto de Kronecker.

Con los estimadores obtenidos a partir de la aplicación de GLS a (2.9) se regresa a (2.8) para obtener un nuevo estimador de $\hat{\beta}$ y se comienza a alternar entre la estimación de los parámetros del componente lineal y la estimación de las covarianzas hasta que el procedimiento converja, es decir, que la estimación de todos los parámetros no cambien de una iteración a otra. Cuando el algoritmo converge se obtienen los estimadores máximo verosímiles.

El procedimiento de estimación por máxima verosimilitud produce de hecho estimadores sesgados de las covarianzas dado que no toma en cuenta la varianza muestral de los parámetros fijos. Esto puede ser importante sobre todo en muestras pequeñas. Se pueden obtener estimadores insesgados por medio de una modificación llamada máxima verosimilitud restringida.

²de *Generalized least squares*

Capítulo 3

Regresión logística multinivel

Ahora se analizará el caso en el que la variable de respuesta es binomial y donde la respuesta puede ser vista como un conteo de eventos. Este conteo puede ser el número de veces que ocurre un evento en un número fijo de realizaciones, en este caso, usualmente se quiere tomar en cuenta la proporción resultante como respuesta.

En el caso del modelo logístico de regresión se quiere estudiar la relación de éxito o fracaso en función de variables de predicción cuando los datos presentan patrones complejos de variabilidad.

3.1. Modelo de intercepto aleatorio sin variables predictoras

En este modelo, el más simple de regresión logística multinivel, se quiere investigar la relación

$$P(Y = 1) = \pi$$

para una variable Y de rango dicotómico cuando las observaciones se encuentran agrupadas en J clases y además se sabe que existe cierta

variación por grupo, es decir, para el grupo j se tiene que

$$P(Y = 1) = \pi_j \quad j = 1, 2, \dots, J.$$

Entonces, por lo visto en el Capítulo 1, se tiene que para el grupo j existe β_{0j} tal que

$$\beta_{0j} = \log \left(\frac{\pi_j}{1 - \pi_j} \right),$$

es el modelo lineal de regresión logística, donde

$$\beta_{0j} \sim \mathcal{N}(\beta_0, \sigma_u^2)$$

o, del mismo modo,

$$\beta_{0j} = \beta_0 + u_j \quad \text{con } u_j \sim \mathcal{N}(0, \sigma_u^2), \quad (3.1)$$

por lo tanto,

$$\pi_j = \frac{e^{\beta_{0j}}}{1 + e^{\beta_{0j}}}$$

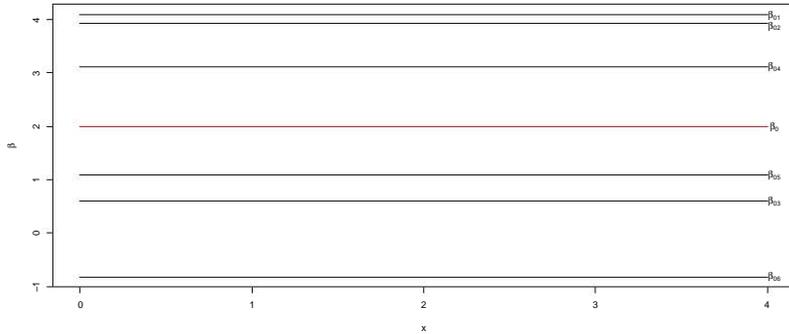
o

$$\pi_j = \frac{e^{\beta_0 + u_j}}{1 + e^{\beta_0 + u_j}}.$$

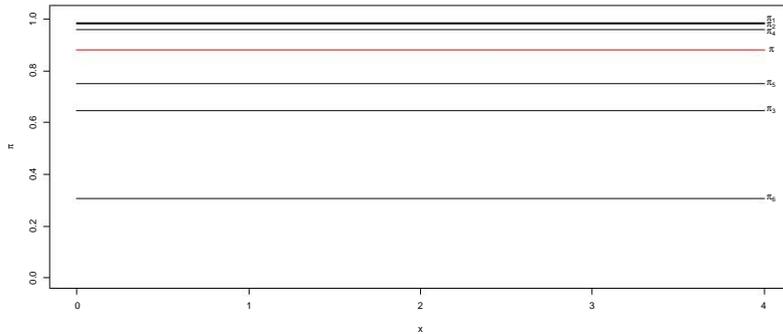
3.1.1. Ejemplo

Supóngase que se tiene una variable Y de tal modo que $\beta_0 = 2$, y $\beta_{0j} = \beta_0 + u_j$. La diferencia en el valor del parámetro ocasionado por las perturbaciones u_j , hacen que la probabilidad de éxito en cada grupo se sea afectada. En el modelo de regresión logística multinivel se hace un ajuste tomando en cuenta estas diferencias para poder hacer una estimación adecuada de β_0 .

En la figura 3.1 se muestra como se comporta la probabilidad de éxito ante la perturbación de β_0 y el reescalamento que se produce.



(a) Valor de β_0 comparado con los β_{0j} .



(b) Valor de π comparado con los π_j .

Figura 3.1: Cambio de escala que se produce al aplicar la inversa de la función logit.

3.2. Modelo de intercepto aleatorio con una variable predictora

Se quiere encontrar la relación

$$P(Y = 1|X = x) = \pi(x).$$

donde el modelo lineal de regresión es

$$\beta_{0j} + \beta_1 x = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right),$$

entonces,

$$\pi(x) = \frac{e^{\beta_{0j} + \beta_1 x}}{1 + e^{\beta_{0j} + \beta_1 x}}$$

donde β_{0j} es como se muestra en la ecuación (3.1), entonces,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x + u_j}}{1 + e^{\beta_0 + \beta_1 x + u_j}}.$$

3.3. Modelo de intercepto y pendiente aleatorios

En este modelo, se supone que se tienen J clases, y para cada clase se cumple que

$$P(Y_j = 1|x) = \pi(x),$$

donde el modelo lineal de regresión es de la forma

$$\beta_{0j} + \beta_{0j} x = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right),$$

con

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \right)$$

o, de manera similar,

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix}$$

con

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \right).$$

Por lo tanto,

$$\frac{e^{\beta_{0j} + \beta_{1j}x}}{1 + e^{\beta_{0j} + \beta_{1j}x}} = \pi(x),$$

entonces,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x + (u_{0j} + u_{1j}x)}}{1 + e^{\beta_0 + \beta_1 x + (u_{0j} + u_{1j}x)}}$$

3.3.1. Ejemplo

Se simularon 475 datos divididos en 5 clases que se muestran en el cuadro 3.1, de tal modo que

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} -5 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.8 & -0.2 \\ -0.2 & 0.5 \end{pmatrix} \right)$$

Entonces, se tiene que

$$P(Y_{ij} = 1|x) = \frac{e^{\beta_{0j} + \beta_{0j}x}}{1 + e^{\beta_{0j} + \beta_{0j}x}}$$

	Muestra 1		Muestra 2		Muestra 3		Muestra 4		Muestra 5	
1	8.8316	1	3.1026	0	2.8918	0	6.8414	0	0.4988	0
2	6.8862	1	1.5623	0	0.3775	0	0.5093	0	3.3451	1
3	2.2538	1	0.2431	0	5.6103	1	4.7717	0	5.7683	1

Continúa en la siguiente página.

	Muestra 1		Muestra 2		Muestra 3		Muestra 4		Muestra 5	
4	7.6348	1	5.979	1	8.0837	1	0.8971	0	5.9126	1
5	0.7216	0	3.4999	1	7.7475	1	5.724	0	7.1787	1
6	1.0203	0	1.7594	0	9.1436	1	1.3908	0	2.235	1
7	9.9673	1	7.5047	1	2.8376	0	9.0407	0	0.6577	0
8	0.5237	0	3.0004	1	9.5223	1	7.1095	1	7.722	1
9	6.0609	0	8.8972	1	2.1338	0	7.4808	0	6.1081	1
10	4.5111	0	8.8102	1	4.1987	1	2.6775	0	0.6055	0
11	0.2224	0	1.9177	0	3.7898	1	2.9611	0	4.0141	1
12	7.6043	1	9.7578	1	8.6583	1	8.9382	0	6.7276	1
13	9.3375	1	7.6403	1	5.3531	1	5.1343	0	2.7619	1
14	3.6467	0	6.3343	1	6.7457	1	8.9463	0	1.0825	0
15	6.1667	0	6.4558	0	1.9885	1	6.217	0	1.1128	0
16	7.6933	1	4.1309	0	5.1749	1	1.1112	0	1.3112	0
17	5.2127	0	3.8963	0	7.3858	1	2.0431	0	3.5409	1
18	7.3027	1	8.5384	1	5.9772	1	7.4578	0	9.7638	1
19	2.9911	0	0.469	0	4.1784	1	7.4616	0	8.4551	1
20	5.1597	0	3.5985	0	6.2271	1	5.3539	0	0.851	1
21	2.5373	0	8.2155	1	1.8322	1	9.0048	0	4.5126	1
22	3.5021	0	5.9002	1	8.1489	1	3.4337	0	3.2911	1
23	3.3425	0	6.2147	1	3.35	1	1.153	0	0.8019	0
24	9.5294	1	3.8582	0	5.4377	1	2.5876	0	9.6665	1
25	1.6575	0	7.9982	1	6.8794	1	2.2386	0	7.843	1
26	4.6798	0	4.8326	0	0.3407	0	4.01	0	1.6197	0
27	2.491	0	4.3445	1	5.1077	1	2.3611	0	7.4988	1
28	4.8325	0	7.5971	1	8.8525	1	5.332	0	1.5732	1
29	2.0983	0	1.1751	0	1.4093	0	2.8503	0	1.1091	1
30	1.7419	0	1.7522	0	0.6095	0	6.5115	0	7.1846	1
31	0.777	0	0.1342	0	1.9737	0	1.1525	0	3.8235	0
32	1.6279	0	3.1705	0	3.4647	1	2.0753	0	4.9128	1
33	1.0259	0	9.3807	1	9.6059	1	5.5326	0	5.6944	1
34	5.2082	0	0.3166	0	1.4416	0	2.772	0	9.9827	1

Continúa en la siguiente página.

	Muestra 1		Muestra 2		Muestra 3		Muestra 4		Muestra 5	
35	0.6358	0	2.4901	0	1.3336	0	9.3324	0	7.1355	1
36	5.3143	0	7.9678	1	6.5662	1	2.1102	0	7.3884	1
37	8.0023	1	4.7142	0	8.1876	1	0.1999	0	0.0208	0
38	9.9573	1	3.2541	0	7.4358	1	4.3563	0	2.1258	0
39	7.629	1	9.7032	1	2.3519	1	5.9035	0	2.8113	1
40	2.8455	0	3.311	0	1.4064	0	4.3714	0	1.7515	0
41	9.0935	1	1.1595	0	3.6124	0	2.7338	0	2.8863	1
42	5.4646	0	2.8699	0	8.0009	1	2.5028	0	3.3896	0
43	1.5619	0	2.8086	0	5.6773	1	2.6248	0	8.4408	1
44	2.7471	0	3.5342	0	4.9248	1	0.4827	0	1.0995	0
45	0.3449	0	5.931	0	4.8967	1	3.7331	0	1.973	0
46	0.0288	0	8.8188	1	3.6527	1	7.7353	0	6.7103	1
47	9.5569	1	7.9733	1	0.9592	0	2.4695	0	1.0016	0
48	7.4398	1	2.8797	1	1.8668	0	7.9706	0	5.8056	1
49	0.7511	0	6.9186	1	1.7414	0	9.239	0	5.4484	1
50	3.5059	0	4.0441	0	8.5098	1	5.3622	0	9.3198	1
51	6.8595	1	9.6162	1	7.6273	1	0.7615	0	0.1574	0
52	5.5065	0	1.7968	0	9.6764	1	0.0528	0	1.4326	0
53	3.2949	0	4.4579	1	8.7088	1	7.6738	0	9.4435	1
54	4.9756	0	2.6219	0	1.1814	0	4.5615	0	4.2837	1
55	6.283	1	5.3295	0	9.3451	1	4.1212	0	2.4445	0
56	0.6139	0	4.1371	0	6.5289	1	1.4538	0	2.4659	1
57	3.1285	1	0.8523	0	8.7713	1	0.7804	0	3.4521	0
58	0.5399	0	4.3305	1	2.2644	0	9.5995	0	9.2641	1
59	9.2266	1	4.3507	1	6.6947	1	7.7388	0	9.0262	1
60	7.212	1	0.8917	0	7.6869	1	7.5418	0	9.0781	1
61	5.6562	0	5.7736	1	7.2585	1	9.6961	0	5.967	1
62	7.3995	1	9.3877	1	1.9308	1	7.7577	0	9.9082	1
63	0.1163	0	5.8715	0	3.8077	1	9.8313	0	0.0566	1
64	1.8934	0	7.9908	1	5.42	1	3.2112	0	6.5951	1
65	8.4952	1	3.1352	0	3.2881	1	6.5742	0	9.8512	1

Continúa en la siguiente página.

	Muestra 1		Muestra 2		Muestra 3		Muestra 4		Muestra 5	
66	4.8276	1	7.7968	1	9.8135	1	1.1084	0	2.7308	1
67	9.0555	1	5.0582	0	1.0084	0	0.4019	0	5.8068	1
68	2.3266	0	2.5597	0	7.3502	1	9.5569	0	4.0242	0
69	6.5253	1	0.5417	0	8.7955	1	4.477	0	1.2346	0
70	3.4229	0	8.5431	1	1.7881	0	5.7843	0	2.6146	1
71	1.7291	0	3.093	1	3.9676	1	0.9457	0	4.7963	1
72	1.0421	0	4.2933	0	6.0698	1	3.0034	0	5.7283	1
73	3.6534	0	1.2083	0	9.8683	1	3.0847	0	3.7006	0
74	4.0957	1	2.315	0	1.402	0	6.2312	0	1.0145	0
75	4.4702	0	5.1937	0	5.6439	1	9.0085	0	0.5043	0
76	4.9781	0	6.9418	1	1.8519	0	3.4433	0	0.3865	0
77	5.2205	0	2.5105	0	3.163	0	9.0215	0	3.6793	1
78	7.4502	1	5.0263	1	4.9123	1	3.6772	0	1.9048	0
79	2.6656	0	2.5227	0	5.3252	1	7.2617	0	3.5531	0
80	6.6995	1	5.1829	1	2.7915	1	1.2457	0	2.5174	1
81	1.5947	0	2.3629	0	7.0698	1	6.0353	0	3.1568	1
82	3.777	0	4.6543	1	4.0861	1	4.419	0	4.4306	1
83	0.0961	0	6.1018	1	1.5134	0	5.865	0	5.3463	1
84	9.0663	1	5.6617	0	1.9197	0	0.4979	0	3.6413	1
85	5.9912	0	7.6129	1	3.0784	1	0.6017	0	4.2752	1
86	9.8621	1	1.8866	0	2.9733	0	4.2079	0	9.2185	1
87	4.2079	1	5.5453	0	4.1473	0	1.8666	0	2.9004	1
88	4.723	0	1.0133	0	4.5201	1	4.2384	0	4.1912	1
89	1.7508	0	1.1899	0	2.757	0	3.9358	0	1.2806	0
90	0.8084	0	8.4906	1	5.2708	1	2.8094	0	1.3921	0
91	0.1132	0	3.852	0	8.152	1	4.0119	0	5.6036	1
92	5.7928	1	8.4557	1	8.7292	1			9.4653	1
93	0.5595	0	9.8694	1					8.7568	1
94	9.2305	1	7.4317	1					9.4101	1
95	1.1432	0	6.7792	1					3.5954	1
96	7.4672	1	0.9626	0					3.908	1

Continúa en la siguiente página.

	Muestra 1		Muestra 2		Muestra 3		Muestra 4		Muestra 5	
97	6.9589	0							3.1758	1
98	2.0291	0								
99	1.197	0								

Cuadro 3.1: *Muestra simulada con pendiente e intercepto aleatorios*

Haciendo la estimación del modelo logístico ignorando la información de la división por clases, se obtienen los estimadores $\widehat{\beta}_{0_{\text{pool}}} \approx -2.5867264$ y $\widehat{\beta}_{1_{\text{pool}}} \approx 0.5017554$. Haciendo la estimación de manera usual por cada clase individualmente, se obtiene:

$$\begin{aligned}
 \widehat{\beta}_{01_{\text{n-pool}}} &= -6.270231 & \widehat{\beta}_{11_{\text{n-pool}}} &= 1.0771042 \\
 \widehat{\beta}_{02_{\text{n-pool}}} &= -5.057761 & \widehat{\beta}_{12_{\text{n-pool}}} &= 1.0164468 \\
 \widehat{\beta}_{03_{\text{n-pool}}} &= -4.801612 & \widehat{\beta}_{13_{\text{n-pool}}} &= 1.6737519 \\
 \widehat{\beta}_{04_{\text{n-pool}}} &= -6.509863 & \widehat{\beta}_{14_{\text{n-pool}}} &= 0.3425247 \\
 \widehat{\beta}_{05_{\text{n-pool}}} &= -2.561314 & \widehat{\beta}_{15_{\text{n-pool}}} &= 1.0831016
 \end{aligned}$$

Por otro lado, haciendo la estimación con el modelo multinivel, se obtienen los estimadores $\widehat{\beta}_{0_{\text{MLR}}} \approx -5.043714$ y $\widehat{\beta}_{1_{\text{MLR}}} \approx 1.020261$. El ajuste que hace este modelo por cada clase es:

$$\begin{aligned}
 \widehat{\beta}_{01_{\text{MLR}}} &= -5.453925 & \widehat{\beta}_{11_{\text{MLR}}} &= 0.9394492 \\
 \widehat{\beta}_{02_{\text{MLR}}} &= -5.060229 & \widehat{\beta}_{12_{\text{MLR}}} &= 1.0170074 \\
 \widehat{\beta}_{03_{\text{MLR}}} &= -3.689285 & \widehat{\beta}_{13_{\text{MLR}}} &= 1.2870837 \\
 \widehat{\beta}_{04_{\text{MLR}}} &= -7.567964 & \widehat{\beta}_{14_{\text{MLR}}} &= 0.5229834 \\
 \widehat{\beta}_{05_{\text{MLR}}} &= -3.347755 & \widehat{\beta}_{15_{\text{MLR}}} &= 1.3543651
 \end{aligned}$$

En la figura 3.2 se muestra la comparativa entre la línea de la simulación, los datos generados y la estimación del modelo por el método *pooling* y el modelo de regresión multinivel (MLR). En la figura 3.3 se presenta una comparativa entre la simulación, los datos generados y las estimaciones

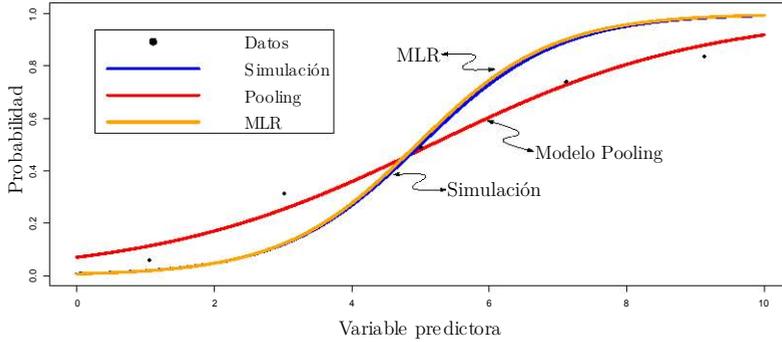


Figura 3.2: *Distintos modelos con coeficientes aleatorios.*

por los modelos usuales. Por último, en la figura 3.4 se presenta una comparativa entre la simulación, los datos generados y la estimación por el modelo logístico multinivel.

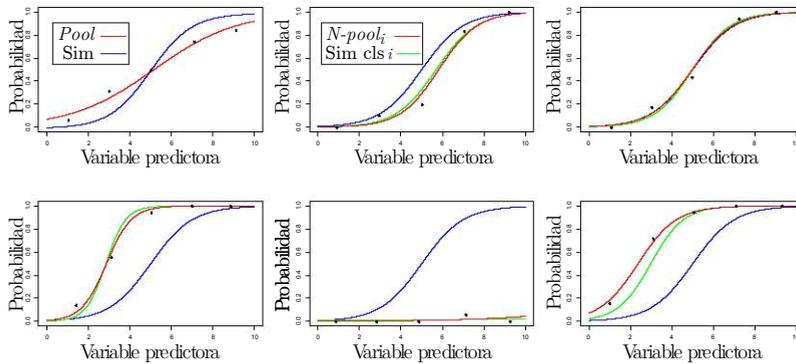


Figura 3.3: *En la primera gráfica se compara el modelo de la simulación, la muestra generada y el modelo pooled estimado. En las siguientes gráficas se compara el modelo general simulado, la clase simulada, los datos pertenecientes a la clase y el modelo usual de regresión de la clase.*

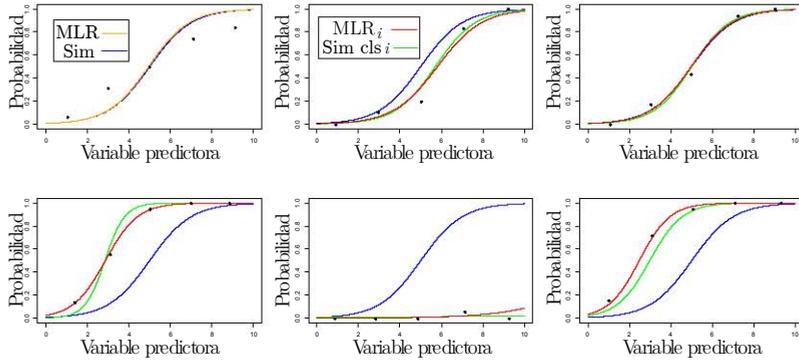


Figura 3.4: En la primera gráfica se compara el modelo de la simulación, la muestra generada y el modelo multinivel estimado. En las siguientes gráficas se compara el modelo general simulado, la clase simulada, los datos pertenecientes a cada clase y el modelo ajustado para la clase dentro del modelo multinivel

		Parámetros de la simulación		Modelos usuales		Modelo multinivel	
Modelo General		β_0	β_1	$\widehat{\beta}_{0\text{pool}}$	$\widehat{\beta}_{1\text{pool}}$	$\widehat{\beta}_{0\text{MLR}}$	$\widehat{\beta}_{1\text{MLR}}$
Clases	1	β_{01}	β_{11}	$\widehat{\beta}_{01\text{n-pool}}$	$\widehat{\beta}_{11\text{n-pool}}$	$\widehat{\beta}_{01\text{MLR}}$	$\widehat{\beta}_{11\text{MLR}}$
	2	β_{02}	β_{12}	$\widehat{\beta}_{02\text{n-pool}}$	$\widehat{\beta}_{12\text{n-pool}}$	$\widehat{\beta}_{02\text{MLR}}$	$\widehat{\beta}_{12\text{MLR}}$
	3	β_{03}	β_{13}	$\widehat{\beta}_{03\text{n-pool}}$	$\widehat{\beta}_{13\text{n-pool}}$	$\widehat{\beta}_{03\text{MLR}}$	$\widehat{\beta}_{13\text{MLR}}$
	4	β_{04}	β_{14}	$\widehat{\beta}_{04\text{n-pool}}$	$\widehat{\beta}_{14\text{n-pool}}$	$\widehat{\beta}_{04\text{MLR}}$	$\widehat{\beta}_{14\text{MLR}}$
	5	β_{05}	β_{15}	$\widehat{\beta}_{05\text{n-pool}}$	$\widehat{\beta}_{15\text{n-pool}}$	$\widehat{\beta}_{05\text{MLR}}$	$\widehat{\beta}_{15\text{MLR}}$

		Parámetros de la simulación		Modelos usuales		Modelo multinivel		Mejor estimación
Modelo General		-5	1	-2.5867264	0.5017554	-5.043714	1.020261	MLR
Clases	1	-5.892828	1.0335511	-6.270231	1.0771042	-5.453925	0.9394492	N-Pool
	2	-5.479526	1.0960095	-5.057761	1.0164468	-5.060229	1.0170074	MLR
	3	-5.948837	2.0932230	-4.801612	1.6737519	-3.689285	1.2870837	N-Pool
	4	-4.914804	0.1099282	-6.509863	0.3425247	-7.567964	0.5229834	N-Pool
	5	-3.793183	1.2742250	-2.561314	1.0831016	-3.347755	1.3543651	MLR

3.4. Estimación de parámetros de la regresión logística multinivel

Los modelos estadísticos para datos con este tipo de estructura pertenecen a los llamados modelos lineales generalizados. Un modelo de dos niveles puede ser escrito de la forma

$$\pi_{ij} = f(X\beta)_{ij}.$$

Donde π_{ij} es el valor esperado de la respuesta para la ij -ésima observación y f es una función no lineal del predictor lineal $(X\beta)_{ij}$. Nótese que se pueden introducir coeficientes aleatorios al segundo nivel. Se debe especificar, además, la distribución de la respuesta $y_{ij}|\pi_{ij}$. Queda entonces especificar la función liga no lineal f , que en este caso es tal que

$$f^{-1} = \log\left(\frac{\pi}{1-\pi}\right).$$

3.5. Proporciones como respuestas

Considerando el modelo de componentes de la varianza para un modelo de dos niveles con una sola variable predictora donde la proporción esperada es modelada usando una función liga logit

$$\pi_{ij} = \frac{e^{X\beta}}{1 + e^{X\beta}} = \left(\frac{1 + e^{X\beta}}{e^{X\beta}}\right)^{-1} = \left(1 + e^{-X\beta}\right)^{-1}.$$

Entonces,

$$\pi_{ij} = \{1 + \exp(-(\beta_0 + \beta_1 x_{ij} + u_{0j}))\}^{-1}.$$

Las respuestas observadas y_{ij} son las proporciones con la hipótesis de que tienen una distribución binomial

$$y_{ij} = \text{Bin}(n_{ij}, \pi_{ij}),$$

donde n_{ij} es el denominador para la proporción. También se tiene

$$\text{Var}(y_{ij}|\pi_{ij}) = \frac{\pi_{ij}(1 - \pi_{ij})}{n_{ij}}.$$

Para enfatizar el vínculo con el modelo cuyas variables de respuesta son de rango continuo, e introducir la estimación por *cuasi verosimilitud*, que consiste en utilizar solamente la media y la varianza de la distribución binomial para obtener la estimación.

Se puede escribir el componente del primer nivel como

$$y_{ij} = \pi_{ij} + \varepsilon_{ij}z_{ij},$$

donde

$$z_{ij} = \sqrt{\frac{\pi_{ij}(1 - \pi_{ij})}{n_{ij}}} \quad \text{y} \quad \sigma_{\varepsilon}^2 = 1.$$

Utilizando esta variable de predicción z y restringiendo la varianza del nivel 1 asociada con su coeficiente aleatorio ε_{ij} a ser 1, se obtiene la varianza binomial requerida en la ecuación anterior. Cuando se ajusta un modelo también se puede estimar la varianza del primer nivel y se puede hacer una comparación con el valor 1.0 obteniendo una prueba de “extra variación binomial”. Esta variación puede surgir por diversas razones.

Si se ha omitido un nivel en el modelo, por ejemplo, si se ha ignorado la agrupación por vivienda en una encuesta con uno o más individuos tomados de de distintas casas, se esperaría que hubiera una variación mayor a la binomial al nivel individual. De la misma manera, supóngase que los individuos y las casas están agrupados en áreas y se escoge agrupar por género y tres grupos de clase social dando por resultado 6 celdas en cada área. Si se manejan estas como unidades del primer nivel para que la respuesta sea una proporción, no se tendrá una variación binomial dado que las proporciones están basadas en la suma de distintas variables binomiales con distintos parámetros. En este caso, la varianza

para el nivel j dentro de un área tendría la forma

$$\frac{\pi_j (1 - \pi_j) - \sigma_1^2}{n_j}$$

donde n_j es el tamaño de la celda. Para ajustar dicho modelo, se tendría que especificar una variable explicativa en el nivel 1 igual a $1/\sqrt{n_j}$ para la j -ésima celda. De manera más general se puede ajustar un modelo con un parámetro extra como el anterior de tal manera que se produzca la siguiente estructura de la varianza del primer nivel (omitiendo los subíndices)

$$\frac{\sigma_0^2 \pi (1 - \pi) + \sigma_1^2}{n}.$$

Por supuesto que no se conoce el valor real de π_{ij} o π_j , de tal modo que en cada iteración se usan estimadores basados en los valores obtenidos en una iteración anterior.

Comentarios finales

Una hipótesis común en los modelos clásicos estadísticos es que las observaciones son independientes e idénticamente distribuidas. En el análisis de regresión, utilizando los modelos lineales usuales, no se puede seguir con la hipótesis de que las observaciones son idénticamente distribuidas dado que estas tendrán distintos valores esperados para valores distintos de las variables predictoras, sin embargo, se continúa con la hipótesis de que los errores son idénticamente distribuidos. Sin embargo, en muchas ocasiones los datos provienen de estructuras más complejas en las cuales estas hipótesis dejan de cumplirse y por lo tanto los modelos usuales pierden validez.

Un problema que se presenta con estas estructuras es la complejidad de integrar esta información en el análisis de regresión. Anteriormente estos problemas habían sido abordados por métodos de agregación o desagregación según fuera conveniente, sin embargo, esto genera el problema de que la solución tenía que diseñarse muy a la medida y con el riesgo de introducir grandes sesgos en la estimación.

Al hacer un análisis de regresión para cada grupo no resulta ser una solución satisfactoria ya que comúnmente las observaciones por cada clase son pocas y las estimaciones de los coeficientes son, por lo tanto, poco estables, además, este tipo de análisis ignora el hecho de que las clases provienen de un mismo proceso y, por tanto, los coeficientes deben

ser, de cierto modo, similares. Esta similitud es utilizada en el análisis logístico multinivel para estabilizar las estimaciones de los grupos con pocas observaciones.

Por otro lado, forzar a los coeficientes de todas las clases a ser iguales es demasiado restrictivo dado que hay muchos factores por los cuales los parámetros de distintas clases podrían ser diferentes. En el análisis logístico multinivel se estiman los coeficientes por cada clase, teniendo en cuenta los valores que las variables predictoras toman en cada uno de los niveles reduciendo el sesgo en la estimación de los parámetros del proceso general.

El uso de la regresión logística multinivel proporciona una manera de estimación intermedia en la cual no se obtiene solamente un conjunto de parámetros estimados y que además no calcula los coeficientes de cada grupo de manera independiente. Esto se realiza tomado en cuenta la hipótesis de parámetros aleatorios, lo cual introduce la necesidad del cálculo de parámetros adicionales para la varianza de estos pero que produce una mejora significativa en la estimación de los parámetros.

Apéndices

Apéndice A

Anexo al Capítulo 1

A.1. Cálculos previos al ajuste de los coeficientes

Se tiene un archivo que contiene los datos del ejemplo guardado en la dirección `C:\Users\PUBLICO\Documents\Hosmerlog.txt`.

ID,	AGE,	AGRP,	CHD																
1,	20,	1,	0	21,	34,	2,	0	41,	41,	4,	0	61,	48,	5,	1	81,	57,	7,	0
2,	23,	1,	0	22,	34,	2,	0	42,	42,	4,	0	62,	48,	5,	1	82,	57,	7,	1
3,	24,	1,	0	23,	34,	2,	1	43,	42,	4,	0	63,	49,	5,	0	83,	57,	7,	1
4,	25,	1,	0	24,	34,	2,	0	44,	42,	4,	0	64,	49,	5,	0	84,	57,	7,	1
5,	25,	1,	1	25,	34,	2,	0	45,	42,	4,	1	65,	49,	5,	1	85,	57,	7,	1
6,	26,	1,	0	26,	35,	3,	0	46,	43,	4,	0	66,	50,	6,	0	86,	58,	7,	0
7,	26,	1,	0	27,	35,	3,	0	47,	43,	4,	0	67,	50,	6,	1	87,	58,	7,	1
8,	28,	1,	0	28,	36,	3,	0	48,	43,	4,	1	68,	51,	6,	0	88,	58,	7,	1
9,	28,	1,	0	29,	36,	3,	1	49,	44,	4,	0	69,	52,	6,	0	89,	59,	7,	1
10,	29,	1,	0	30,	36,	3,	0	50,	44,	4,	0	70,	52,	6,	1	90,	59,	7,	1
11,	30,	2,	0	31,	37,	3,	0	51,	44,	4,	1	71,	53,	6,	1	91,	60,	8,	0
12,	30,	2,	0	32,	37,	3,	1	52,	44,	4,	1	72,	53,	6,	1	92,	60,	8,	1
13,	30,	2,	0	33,	37,	3,	0	53,	45,	5,	0	73,	54,	6,	1	93,	61,	8,	1
14,	30,	2,	0	34,	38,	3,	0	54,	45,	5,	1	74,	55,	7,	0	94,	62,	8,	1
15,	30,	2,	0	35,	38,	3,	0	55,	46,	5,	0	75,	55,	7,	1	95,	62,	8,	1
16,	30,	2,	1	36,	39,	3,	0	56,	46,	5,	1	76,	55,	7,	1	96,	63,	8,	1
17,	32,	2,	0	37,	39,	3,	1	57,	47,	5,	0	77,	56,	7,	1	97,	64,	8,	0
18,	32,	2,	0	38,	40,	4,	0	58,	47,	5,	0	78,	56,	7,	1	98,	64,	8,	1
19,	33,	2,	0	39,	40,	4,	1	59,	47,	5,	1	79,	56,	7,	1	99,	65,	8,	1
20,	33,	2,	0	40,	41,	4,	0	60,	48,	5,	0	80,	57,	7,	0	100,	69,	8,	1

Primero se cargarán los datos a R , utilizando las siguientes instrucciones.

```
_____ Código de R _____  
cuadro1<-read.table(  
_____
```

```
"C:/Users/PUBLICICO/Documents/Hosmerlog.txt",
  sep=",", header=TRUE)
attach(cuadro1) _____ Código de R _____
```

Para construir el cuadro 1.2, se utilizan las siguientes instrucciones.

```
_____ Código de R _____
nn<-length(unique(AGRP))
grupol<-NULL
grupou<-NULL
n<-NULL
Ausente<-NULL
Presente<-NULL
for(i in 1:nn){
  grupol<-c(grupol,min(AGE[AGRP==i]))
  grupou<-c(grupou,max(AGE[AGRP==i]))
  n<-c(n,length(AGE[AGRP==i]))
  Presente<-c(Presente,sum(CHD[AGRP==i]))
}
Media<-Presente/n
Ausente<-n-Presente
aux<-cbind(grupol,
           grupou,n,Presente,Ausente,Media)
cuadro2<-rbind(aux,
               c(mean(AGE),mean(AGE),sum(n),
                 sum(Ausente),sum(Presente),mean(Media)))
rownames(cuadro2)<-c(rep(" ",nn),"Total")
cuadro2 _____ Código de R _____
```

```
_____ R output _____
```

grupol	grupou	n	Presente	Ausente	Media
20.00	29.00	10	1	9	0.1000000
30.00	34.00	15	2	13	0.1333333
35.00	39.00	12	3	9	0.2500000
40.00	44.00	15	5	10	0.3333333

	45.00	49.00	13	6	7	0.4615385
	50.00	54.00	8	5	3	0.6250000
	55.00	59.00	17	13	4	0.7647059
	60.00	69.00	10	8	2	0.8000000
Total	44.38	44.38	100	57	43	0.4334889

R output

A.2. Ajuste de los coeficientes de regresión logística con el uso de *R*

Para hacer las estimaciones utilizando *R*, se usará la función `glm` que pertenece a la biblioteca `stats` y que viene cargada en la instalación básica de *R*.

La función `glm` se utiliza para ajustar modelos lineales generalizados. Recibe una representación simbólica del predictor lineal y la distribución de los errores del modelo.

```

Código de R
> (L<-glm(CHD~AGE, family=binomial("logit")))
Código de R
R output

Call:  glm(formula = CHD ~ AGE,
           family = binomial("logit"))

Coefficients:
(Intercept)          AGE
   -5.3095         0.1109

Degrees of Freedom: 99 Total (i.e. Null);
                   98 Residual

Null Deviance:      136.7
Residual Deviance: 107.4      AIC: 111.4
R output

```

Y, para obtener la información como aparece en el cuadro 1.3, se utilizan las siguientes instrucciones.

```

_____ Código de R _____
(cuadro3<-summary(L)$coef)
_____ Código de R _____

```

```

_____ R output _____
      Estimate Std.Error z value Pr(>|z|)
(Intercept) -5.309453  1.133653 -4.68348 2.8203e-06
AGE          0.110921  0.024059  4.61022 4.0223e-06
_____ R output _____

```

Apéndice B

Anexo al Capítulo 2

B.1. Ajuste de los coeficientes aleatorios con el uso de *R*

Para hacer las estimaciones utilizando *R*, se utilizarán las bibliotecas *arm* y *MASS*, por lo que se deben incluir utilizando los siguientes comandos:

<pre>library (arm) library (MASS)</pre>	Código de <i>R</i>
---	--------------------

B.1.1. Modelo de intercepto aleatorio sin variables predictoras

En este caso, el modelo es:

$$y_{ij} = \beta + u_j + \varepsilon_{ij}$$

con $i_j = 1, 2, \dots, n_j$, donde $\beta \in \mathbb{R}$, $u_j \sim \mathcal{N}(0, \sigma_u^2)$ y $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Se hará el ajuste de los datos contenidos en el cuadro 2.1 utilizando *R*. Para esto supondremos que tenemos los datos de la columna *X* del cuadro 2.1 en una variable de tipo `vector` llamada `x` y, en una variable del mismo tipo llamada `cls`, los datos de la columna *cls*.

B.1.1.1. Modelo *complete-pooling*

Para hacer el análisis ignorando la información de división por clases, al no tener variables predictoras, simplemente se hace el promedio de todos los elementos de la variable X de la siguiente manera

```

Código de R
(b0.pooled<-mean(X))
Código de R
R output
[1] 3.550418
R output

```

B.1.1.2. Modelo *no-pooling*

Ahora se tiene que ajustar el modelo ajustando un estimador por cada clase. Es importante notar que `cls` ya está codificada de manera conveniente.

```

Código de R
b0.no.pooled<-NULL
for(i in 1:length(table(cls))) {
b0.no.pooled[i]<-mean(X[cls==i]) }
b0.no.pooled
Código de R
R output
[1] 1.702043 5.611833 3.672684 7.690917 8.508533
[6] 0.883950 6.292571 1.576000 2.662200 2.106467
R output

```

B.1.1.3. Modelo multinivel

B.1.1.3.1. Estimación con máxima verosimilitud

```

Código de R
glmmPQL(X~1, random=~1|cls, family=gaussian)
Código de R
R output
iteration 1
iteration 2
Linear mixed-effects model fit by

```

```

maximum likelihood
  Data: NULL
  Log-likelihood: NA
  Fixed: X ~ 1
(Intercept)
  3.943733

Random effects:
Formula: ~1 | cls
          (Intercept) Residual
StdDev:    2.296304 4.725353

Variance function:
Structure: fixed weights
Formula: ~invwt
Number of Observations: 158
Number of Groups: 10

```

R output

B.1.1.3.2. Estimación con máxima verosimilitud restringida

```

Código de R
lmer(formula=X~1+(1|cls))
Código de R
R output
Linear mixed model fit by REML
Formula: X ~ 1 + (1 | cls)
   AIC   BIC logLik deviance REMLdev
 958.9 968.1 -476.4   954.4   952.9
Random effects:
Groups   Name             Variance Std.Dev.
cls      (Intercept)    6.033   2.4562
Residual                   22.328   4.7253
Number of obs: 158, groups: cls, 10

Fixed effects:

```

	Estimate	Std. Error	t value
(Intercept)	3.9559	0.8697	4.549

R output

B.1.2. Intercepto aleatorio con una variable predictora

El modelo es

$$y_{ij} = \beta_0 + u_j + \beta_1 x$$

donde $u_j \sim \mathcal{N}(\beta_0, \sigma_u^2)$.

Se hará el ajuste de los datos contenidos en el cuadro 2.2 utilizando *R*. Se tienen los datos de la columna *Y* del cuadro 2.2 en una variable de tipo vector llamada *Y*, los datos de la columna *X* se tienen en una variable del mismo tipo llamada *X* y los datos de la columna *cls* en la variable *cls*

B.1.2.1. Modelo *complete-pooling*

Para hacer el análisis ignorando la información de división por clases, al no tener variables predictoras, simplemente se hace el promedio de todos los elementos de la variable *X* de la siguiente manera

```

Código de R
lm(Y~X)
Código de R
R output
Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)          X
  8.790640       3.559857
R output

```

B.1.2.2. Modelo *no-pooling*

Ahora se tiene que ajustar el modelo ajustando un estimador por cada clase. Es importante notar que *cls* ya está codificada de manera

conveniente.

```

_____ Código de R _____
b.no.pooled<-NULL
for(i in 1:length(table(cls))){
b.no.pooled<-
  rbind(b.no.pooled,lm(Y[cls==i]~X[cls==i])
        $coefficients)}
b.no.pooled
_____ Código de R _____

```

```

_____ R output _____
      (Intercept) X[cls == i]
[1,] -26.905531    3.953232
[2,]  -3.515473    4.226143
[3,]   1.823689    3.885643
[4,] -13.188425    3.806271
[5,]   3.931460    3.729360
[6,]  18.470271    4.225655
[7,]  32.790093    4.161934
[8,]  10.255675    4.054525
[9,]  15.287811    3.595181
[10,]  8.636717    3.917612
_____ R output _____

```

B.1.2.3. Modelo multinivel

B.1.2.3.1. Estimación con máxima verosimilitud

```

_____ Código de R _____
glmmPQL(Y~X,random=~1|cls,family=gaussian)
_____ Código de R _____

```

```

_____ R output _____
iteration 1
iteration 2
Linear mixed-effects model fit by
maximum likelihood
  Data: NULL
  Log-likelihood: NA
_____ R output _____

```

```

Fixed: Y ~ X
(Intercept)          X
      3.951410      4.044657

Random effects:
Formula: ~1 | cls
      (Intercept) Residual
StdDev:   16.16133  1.877845

Variance function:
Structure: fixed weights
Formula: ~invwt
Number of Observations: 65
Number of Groups: 10

```

R output

B.1.2.3.2. Estimación con máxima verosimilitud restringida

```

Código de R
lmer(formula=Y~X+(1|cls))
Código de R
R output
Linear mixed model fit by REML
Formula: Y ~ X + (1 | cls)
      AIC   BIC logLik deviance REMLdev
  333.7 342.4 -162.9   327.1   325.7
Random effects:
Groups   Name          Variance Std.Dev.
cls      (Intercept) 290.2527 17.0368
Residual                3.5917  1.8952
Number of obs: 65, groups: cls, 10

Fixed effects:
              Estimate Std. Error t value
(Intercept)  3.95120    5.41531    0.73
X             4.04470    0.05962   67.84

```

```
Correlation of Fixed Effects:
  (Intr)
X -0.088
```

R output

B.1.3. Intercepto y pendiente aleatorios con una variable predictora

El modelo es

$$y_{ij} = \beta_0 + \beta_1 x + (u_{0j} + u_{1j}x + \varepsilon_{ij})$$

Donde

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} \\ \rho_{u_0 u_1} \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \right)$$

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

son independientes.

Se hará el ajuste de los datos contenidos en el cuadro 2.3 utilizando R. Se tienen los datos de la columna X del cuadro 2.3 en una variable de tipo vector llamada X, los datos de la columna Y se tienen en una variable del mismo tipo llamada Y y los datos de la columna cls en la variable cls.

B.1.3.1. Modelo *complete-pooling*

Para hacer el análisis ignorando la información de división por clases, al no tener variables predictoras, simplemente se hace el promedio de todos los elementos de la variable X de la siguiente manera

```
lm(Y~X)
Código de R

lm(formula = Y ~ X)
Call:
lm(formula = Y ~ X)
R output
```

```

Coefficients:
(Intercept)          X
 17.939408      1.000137
R output

```

B.1.3.2. Modelo *no-pooling*

Ahora se tiene que ajustar el modelo ajustando un estimador por cada clase. Es importante notar que `cls` ya está codificada de manera conveniente.

```

Código de R
b.no.pooled<-NULL
for(i in 1:length(table(cls))){
b.no.pooled<-
  rbind(b.no.pooled,lm(Y[cls==i]~X[cls==i])
          $coefficients)}
b.no.pooled
Código de R

```

```

R output
(Intercept) X[cls == i]
[1,] -7.382533 7.356191
[2,] 10.522905 9.620880
[3,] -5.154519 -9.494234
[4,] 23.782920 -2.675796
[5,] 7.333373 1.948352
[6,] -10.442093 7.275470
[7,] 7.841288 2.868004
[8,] 9.438124 11.205902
[9,] 10.128989 1.290212
[10,] -6.434780 10.352460
R output

```

B.1.3.3. Modelo multinivel

B.1.3.3.1. Estimación con máxima verosimilitud

```

Código de R
-----
glmmpQL(Y~X, random=~X|cls, family=gaussian)
Código de R
-----
R output

```

```

iteration 1
iteration 2
Linear mixed-effects model fit by
maximum likelihood
  Data: NULL
  Log-likelihood: NA
  Fixed: Y ~ X
(Intercept)                X
    3.981196      3.972071

Random effects:
Formula: ~X | cls
Structure: General positive-definite,
           Log-Cholesky parametrization
           StdDev   Corr
(Intercept) 10.146541 (Intr)
X           6.202308 -0.171
Residual    1.731029

Variance function:
Structure: fixed weights
Formula: ~invwt
Number of Observations: 65
Number of Groups: 10
R output

```

B.1.3.3.2. Estimación con máxima verosimilitud restringida

```

Código de R
-----
lmer(Y~X+(1+X|cls))
Código de R
-----

```

```
----- R output -----  
Linear mixed model fit by REML  
Formula: Y ~ X + (1 + X | cls)  
   AIC   BIC logLik deviance REMLdev  
 383.9 396.9 -185.9   379.4   371.9  
Random effects:  
Groups   Name          Variance Std.Dev. Corr  
cls      (Intercept) 114.7052 10.7101  
         X           42.7507  6.5384 -0.171  
Residual                2.9965  1.7310  
Number of obs: 65, groups: cls, 10  
  
Fixed effects:  
              Estimate Std. Error t value  
(Intercept)    3.979      3.428   1.161  
X              3.972      2.069   1.920  
  
Correlation of Fixed Effects:  
 (Intr)  
X -0.173  
----- R output -----
```

Apéndice C

Producto Kronecker

El producto de Kronecker, producto directo o producto tensorial, es una operación utilizada en el estudio de las ecuaciones matriciales y en diversas aplicaciones tales como el cálculo de mínimos cuadrados.

Este producto está definido para dos matrices de tamaño arbitrario con valores en cualquier anillo, pero usualmente es de interés cuando las matrices tienen entradas en un campo \mathbb{F} .

Definición. El *producto Kronecker* de dos matrices A y B tales que $A = [a_{ij}] \in M_{m \times n}(\mathbb{F})$, $B \in M_{p \times q}(\mathbb{F})$, denotado por $A \otimes B$ se define como la matriz en bloques

$$A \otimes B \equiv \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \in M_{mp \times nq}(\mathbb{F}).$$

Nótese que, en general, $A \otimes B \neq B \otimes A$.

Propiedad C.1. Si $x, y \in \mathbb{F}^n$, se cumple que $xy' = x \otimes y'$

Demostración. Sea $x = (x_1, x_2, \dots, x_n)$ y $y = (y_1, y_2, \dots, y_n)$, entonces,

$$\begin{aligned} xy' &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}' = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1 \ y_2 \ \cdots \ y_n) \\ &= \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{pmatrix} = \begin{pmatrix} x_1 y' \\ x_2 y' \\ \vdots \\ x_n y' \end{pmatrix} \\ &= x \otimes y' \end{aligned}$$

□

Definición. Sea $A \in M_{m \times n}(\mathbb{F})$. La k -ésima potencia Kronecker, denotada por $A^{\otimes k}$, se define inductivamente para todo $k \in \mathbb{N}^+$ como $A^{\otimes 1} \equiv A$ y

$$A^{\otimes k} \equiv A \otimes A^{\otimes(k-1)}, \quad k = 2, 3, \dots$$

Propiedad C.2. Si $A \in M_{m \times n}(\mathbb{F})$, entonces $A^{\otimes k} \in M_{m^k \times n^k}(\mathbb{F})$

Demostración. Por inducción, sea $A \in M_{m \times n}(\mathbb{F})$, entonces,

$$A^{\otimes 1} = A \in M_{m \times n}(\mathbb{F}).$$

Se supone que se cumple para $k-1$, es decir, $A^{\otimes(k-1)} \in M_{m^{(k-1)} \times n^{(k-1)}}(\mathbb{F})$, entonces,

$$A^{\otimes k} = A \otimes A^{\otimes(k-1)} \in M_{m m^{(k-1)} \times n n^{(k-1)}}(\mathbb{F}) = M_{m^k \times n^k}(\mathbb{F}).$$

□

Propiedad C.3. El producto Kronecker de $I \in M_m(\mathbb{F})$ con $I \in M_n(\mathbb{F})$ es $I \in M_{mn}(\mathbb{F})$.

Demostración.

$$I_m \otimes I_n = \begin{pmatrix} 1I_n & 0I_n & \cdots & 0I_n \\ 0I_n & 1I_n & \cdots & 0I_n \\ \vdots & \vdots & \ddots & \vdots \\ 0I_n & 0I_n & \cdots & 1I_n \end{pmatrix} = \begin{pmatrix} I_n & 0 & \cdots & 0 \\ 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_n \end{pmatrix} = I_{mn}.$$

□

Propiedad C.4. Sea $A \in M_{m \times n}(\mathbb{F})$, entonces, siempre se cumple que $A \otimes 0 = 0 \otimes A = 0 \in M_{m \times n}(\mathbb{F})$.

Demostración.

$$A \otimes 0 = [a_{ij}0] = [0] = 0 \in M_{m \times n}(\mathbb{F})$$

Para $i = 1, 2, \dots, m; j = 1, 2, \dots, n$.

Por otro lado,

$$0 \otimes A = [0A] = 0A = 0 \in M_{m \times n}(\mathbb{F}).$$

Por lo tanto, $A \otimes 0 = 0 \otimes A = 0 \in M_{m \times n}(\mathbb{F})$.

□

Propiedad C.5. $(\alpha A) \otimes B = A \otimes (\alpha B)$ para toda $\alpha \in \mathbb{F}$, $A \in M_{m \times n}(\mathbb{F})$, $B \in M_{p \times q}(\mathbb{F})$.

Demostración.

$$(\alpha A) \otimes B = [(\alpha a_{ij}) B] = [a_{ij} (\alpha B)] = A \otimes (\alpha B)$$

Para $i = 1, 2, \dots, m; j = 1, 2, \dots, n$.

□

Propiedad C.6. $(A \otimes B)' = A' \otimes B'$ para $A \in M_{m \times n}(\mathbb{F})$ y $B \in M_{p \times q}(\mathbb{F})$.

Demostración.

$$\begin{aligned}
 (A \otimes B)' &= \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}' \\
 &= \begin{pmatrix} a_{11}B' & a_{21}B' & \cdots & a_{m1}B' \\ a_{12}B' & a_{22}B' & \cdots & a_{m2}B' \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}B' & a_{2n}B' & \cdots & a_{mn}B' \end{pmatrix} \\
 &= A' \otimes B'.
 \end{aligned}$$

□

Propiedad C.7. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$, para $A \in M_{m \times n}(\mathbb{F})$, $B \in M_{p \times q}(\mathbb{F})$ y $C \in M_{r \times s}(\mathbb{F})$.

Demostración.

$$\begin{aligned}
 (A \otimes B) \otimes C &= \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} \otimes C \\
 &= \begin{pmatrix} (a_{11}B) \otimes C & (a_{12}B) \otimes C & \cdots & (a_{1n}B) \otimes C \\ (a_{21}B) \otimes C & (a_{22}B) \otimes C & \cdots & (a_{2n}B) \otimes C \\ \vdots & \vdots & \ddots & \vdots \\ (a_{m1}B) \otimes C & (a_{m2}B) \otimes C & \cdots & (a_{mn}B) \otimes C \end{pmatrix} \\
 &= \begin{pmatrix} a_{11}(B \otimes C) & a_{12}(B \otimes C) & \cdots & a_{1n}(B \otimes C) \\ a_{21}(B \otimes C) & a_{22}(B \otimes C) & \cdots & a_{2n}(B \otimes C) \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}(B \otimes C) & a_{m2}(B \otimes C) & \cdots & a_{mn}(B \otimes C) \end{pmatrix} \\
 &= A \otimes (B \otimes C).
 \end{aligned}$$

□

Propiedad C.8. $(A + B) \otimes C = A \otimes C + B \otimes C$, para $A, B \in M_{m \times n}(\mathbb{F})$ y $C \in M_{p \times q}(\mathbb{F})$.

Demostración.

$$\begin{aligned} (A + B) \otimes C &= [(a_{ij} + b_{ij}) C] = [a_{ij} C + b_{ij} C] \\ &= [a_{ij} C] + [b_{ij} C] = A \otimes C + B \otimes C. \end{aligned}$$

□

Propiedad C.9. $A \otimes (B + C) = A \otimes B + A \otimes C$, para $A \in M_{m \times n}(\mathbb{F})$, y $B, C \in M_{p \times q}(\mathbb{F})$.

Demostración.

$$\begin{aligned} A \otimes (B + C) &= [a_{ij} (B + C)] = [(a_{ij} B) + (a_{ij} C)] = \\ &= [a_{ij} B] + [a_{ij} C] = A \otimes B + A \otimes C. \end{aligned}$$

Para $i = 1, 2, \dots, m; j = 1, 2, \dots, n$.

□

Propiedad C.10. $A \otimes B = 0$ si y solo si $A = 0$ ó $B = 0$.

Demostración. P.D. Si $A = 0$ ó $B = 0$ entonces $A \otimes B = 0$.

Supóngase que $A = 0$ ó $B = 0$, entonces, por la propiedad C.4, se tiene que $A \otimes B = 0$. ■

P.D. Si $A \otimes B = 0$ entonces $A = 0$ ó $B = 0$. Ahora, supóngase que $A \otimes B = 0$, entonces

$$\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Entonces, se cumple que para toda $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$, $B = 0$ ó $a_{ij} = 0$. Por lo tanto, $A = 0$ ó $B = 0$. □

Índice alfabético

- Agregación, 29
- Análisis multinivel, 27
- Cociente de verosimilitudes, 19
- Coefficiente de correlación intraclassa, 51
- Cuasi verosimilitud, 69
- Desagregación, 28, 29
- Devianza, 19
- Ecuaciones de verosimilitud, 13
- Error estándar estimado, 23
- Estimación
 - logística, 13
 - logística multinivel, 68
 - multinivel, 50
- Estimador máximo verosímil, 8, 55
- Familia exponencial, 6
 - verosimilitud, 7
- Función logit, 15
- Kronecker
 - potencia, 90
 - producto, 89
- Liga canónica, 7
- Logit, 15, 68
- Matriz de información observada, 22
- Modelo
 - de intercepto aleatorio, 35
 - de intercepto y pendientes aleatorias, 36
 - de pendientes aleatorias, 36
 - jerárquico, 27
- Modelo de regresión lineal, 14
- Modelo lineal de regresión logística, 7, 58
- Producto
 - directo, 89
 - Kronecker, 55, 89
 - tensorial, 89
- Prueba
 - F , 17
 - t , 18

de Wald univariada, 24

Regresión logística, 5

 multinivel, 57

 caracterización del modelo, 15

Suma de cuadrados

 de la regresión, 17

 de residuales, 17

 totales, 17

Tabla de análisis de varianza, 17

Variables

dummy, 32

mudas, 32

 dicotómicas, 5

 significativas, 16

Bibliografía

Jan de Leeuw and Erik Meijer, *Handbook of multilevel analysis*, vol. 1, Springer, ene 2008.

Andrew Gelman and Jennifer Hill, *Data analysis using regression and multilevel/hierarchical models*, vol. 1, Cambridge University Press, dic 2006.

Harvey Goldstein, *Multilevel statistical models*, vol. 1, Wiley, nov 2010.

Roger A. Horn, *Topics in matrix analysis*, vol. 1, Cambridge University Press, jun 1994.

David W. Hosmer Jr., Stanley Lemeshow, and Rodney X. Sturdivant, *Applied logistic regression*, vol. 1, Wiley, abr 2013.

Hasinur Rahaman Khan and J Ewart H Shaw, *Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data*, 93–110.

Jeffrey S. Simonoff, *Analyzing categorical data*, vol. 1, Springer, jul 2003.

Jos W. R. Twisk, *Applied multilevel analysis: A practical guide for medical researchers*, vol. 1, Cambridge University Press, mar 2006.