REGRESIÓN LOGÍSTICA

Introducción

- En ocasiones analizamos datos donde la variable dependiente de interés toma valores discretos:
  - Variable dependientes binarias: endeudarse o no
  - Variables discretas sin ordenación: modo de transporte (tren, autobús, etc…)
  - Variables discretas con orden: calificación/rating financiero
- Usamos la regresión logística cuando el fenómeno estudiado tiene una variable de respuesta cualitativa.

ESTUDIOS EPIDEMIOLÓGICOS

- Estudios relacionados con enfermedades (cáncer, diabetes, hipertensión, ..)
- o eventos (fallecimiento en cirugías)
  - VARIABLE DEPENDIENTE

$$Y_i = \begin{cases} 1 & si \quad hay \quad enfermedad \\ 0 & no \quad hay \quad enfermedad \end{cases}$$

  - VARIABLES Y FACTORES EXPLICATIVAS
    - Edad
    - Género
    - Fumador
    - Antecedentes familiares
    - Otras enfermedades
    - Alimentación
    - Duración cirugía
    - Medicamento

ACCIDENTES DE TRÁFICO MORTALES EN COLOMBIA - FACTORES DE RIESGO

- VARIABLE DEPENDIENTE

$$Y_i = \begin{cases} 1 & accidente \quad con \quad muerte \\ 0 & accidente \quad \sin \quad muerte \end{cases}$$

- VARIABLES Y FACTORES EXPLICATIVAS
  - Horas del día
  - Día de la semana
  - Condiciones meteorológicas: lluvia, niebla, viento
  - Variaciones geográficas: área urbana, área rural,
  - Ciudades Bogotá, Medellín, Cali, Barranquilla, Bucaramanga.
  - Sector: residencial, comercial, industrial (en ciudades)
  - Zona: escolar, no escolar (en ciudades)
  - Tipo de vehículo
  - Número de pasajeros
  - Víctima conductor

DESERCIÓN ESTUDIANTIL PRIMARIA, GRADOS 1 A 5

- VARIABLE DEPENDIENTE

$$Y_i = \begin{cases} 1 & estudiante \quad deserta \\ 0 & estudiante \quad no \quad deserta \end{cases}$$

- VARIABLES Y FACTORES EXPLICATIVAS

Variables extra-escolares	Variables intra-escolares
Características del niño:	**Características de la institución educativa**:
-Género	-Tipo de jornada
-Edad (extra-edad, sub-edad)	-Metodología
-Pertenencia a una etnia	-Número de docentes universitarios
-Victima del conflicto	-Presencia de laboratorios
-Presencia de alguna discapacidad	-Presencia de salas de computadores
-Proveniencia de otro municipio	-Presencia de biblioteca
-Proveniencia del sector privado	-Presencia de zonas verdes en buen estado
-Repitencia de años pasados	**Clima escolar**:
Características del hogar:	-Número de alumnos por salón
-Educación del jefe del hogar	-Problemas de hacinamiento
-Estrato socio-económico	-Existencia de un manual de convivencia
-Ingresos mensuales	-Proyectos de uso del tiempo libre

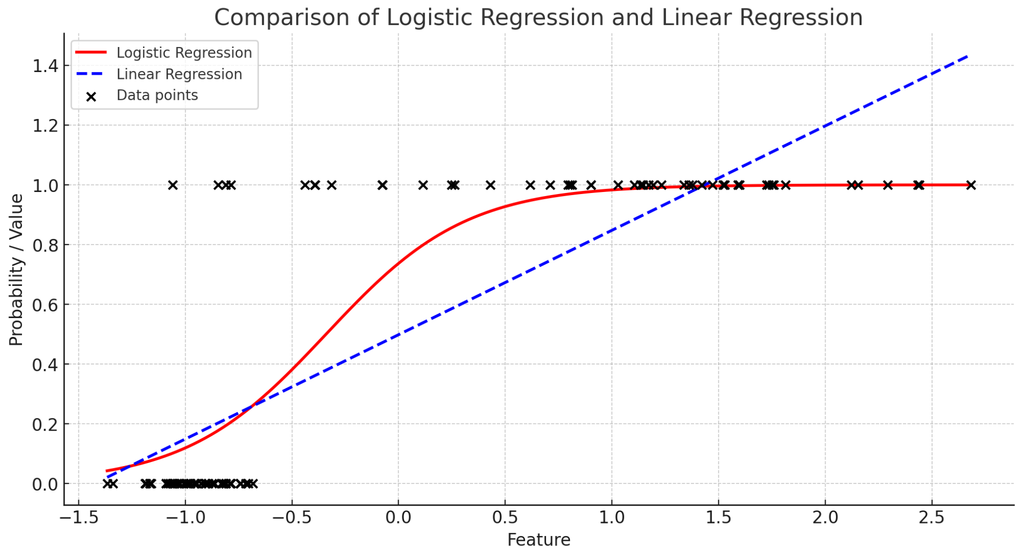
Modelo Lineal de Probabilidad

- El Modelo Lineal de Probabilidad simplemente supone que la esperanza condicional de la variable binaria Y es lineal.

$$E\left(Y \middle| X = x\right) = P\left(Y = 1 \middle| X = x\right) = p\left(x\right) = \beta_0 + \beta_1 x$$

- Todo lo que ya sabemos sobre el modelo de regresión lineal se puede aplicar directamente: estimación, contraste de hipótesis, interpretación de los parámetros, etc.
- Sólo debemos recordar que la esperanza condicional es, en este caso, una probabilidad.
- ¿Por qué no se utiliza frecuentemente el Modelo Lineal de Probabilidad?

Se quiere relacionar Y con $X1, X2, \ldots, Xk$, pero ahora la variable dependiente Y es una variable dicotómica que toma únicamente dos valores



- Para Y continua un modelo lineal puede ajustar bien los datos. El estimador de mínimos cuadrados tiene las mejores propiedades estadísticas
- Para Y dicotómica, un modelo lineal no es apropiado y el estimador por mínimos cuadrados deja de tener buenas propiedades estadísticas

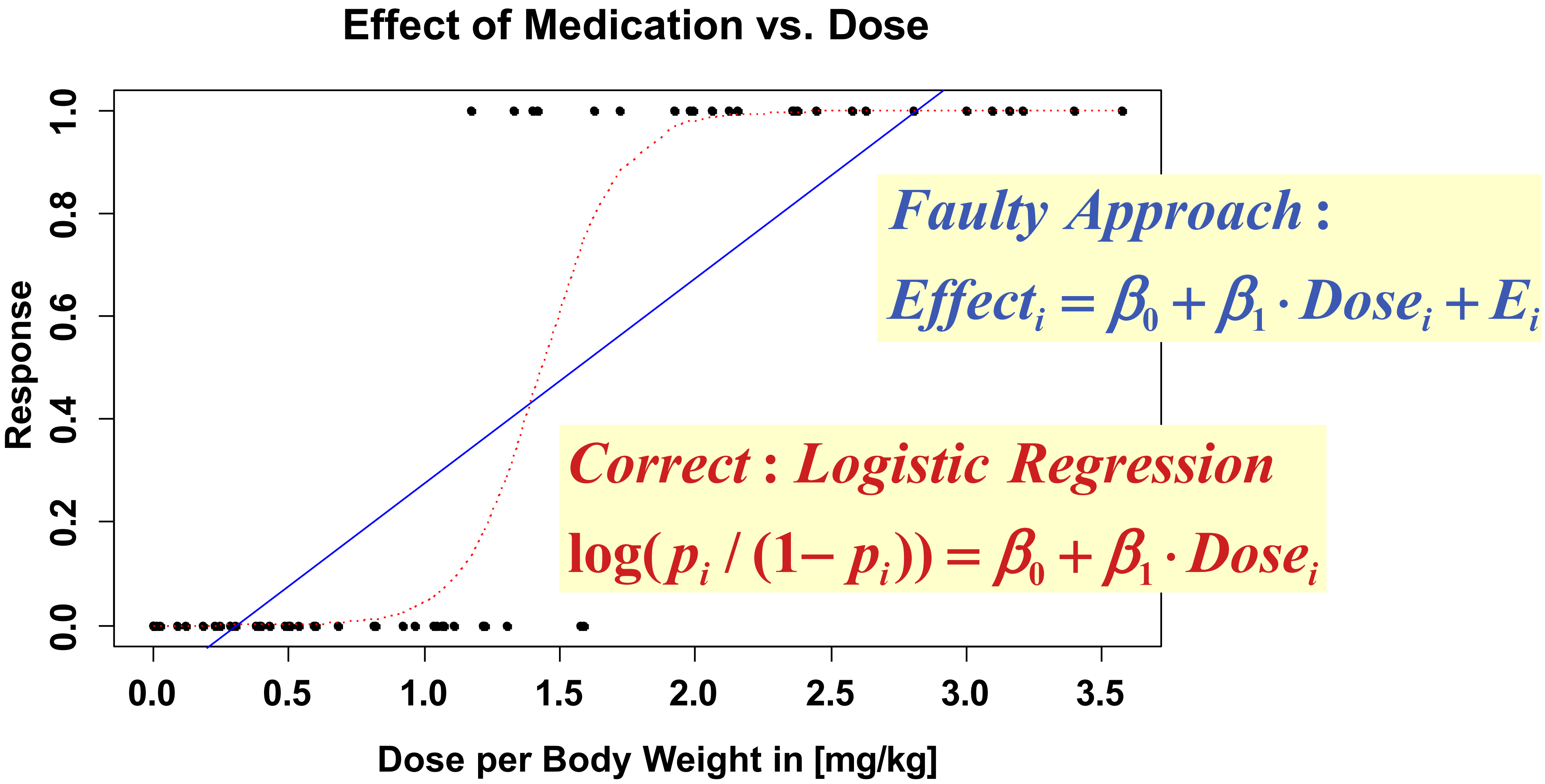
MLP: limitaciones

- Heterocedasticidad. El Modelo Lineal de Probabilidad es, por construcción, heterocedástico

$$Var\left(Y \middle| X = x\right) = p\left(x\right)\left[1 - p\left(x\right)\right] = \left(\beta_0 + \beta_1 x\right)\left(1 - \beta_0 - \beta_1 x\right)$$

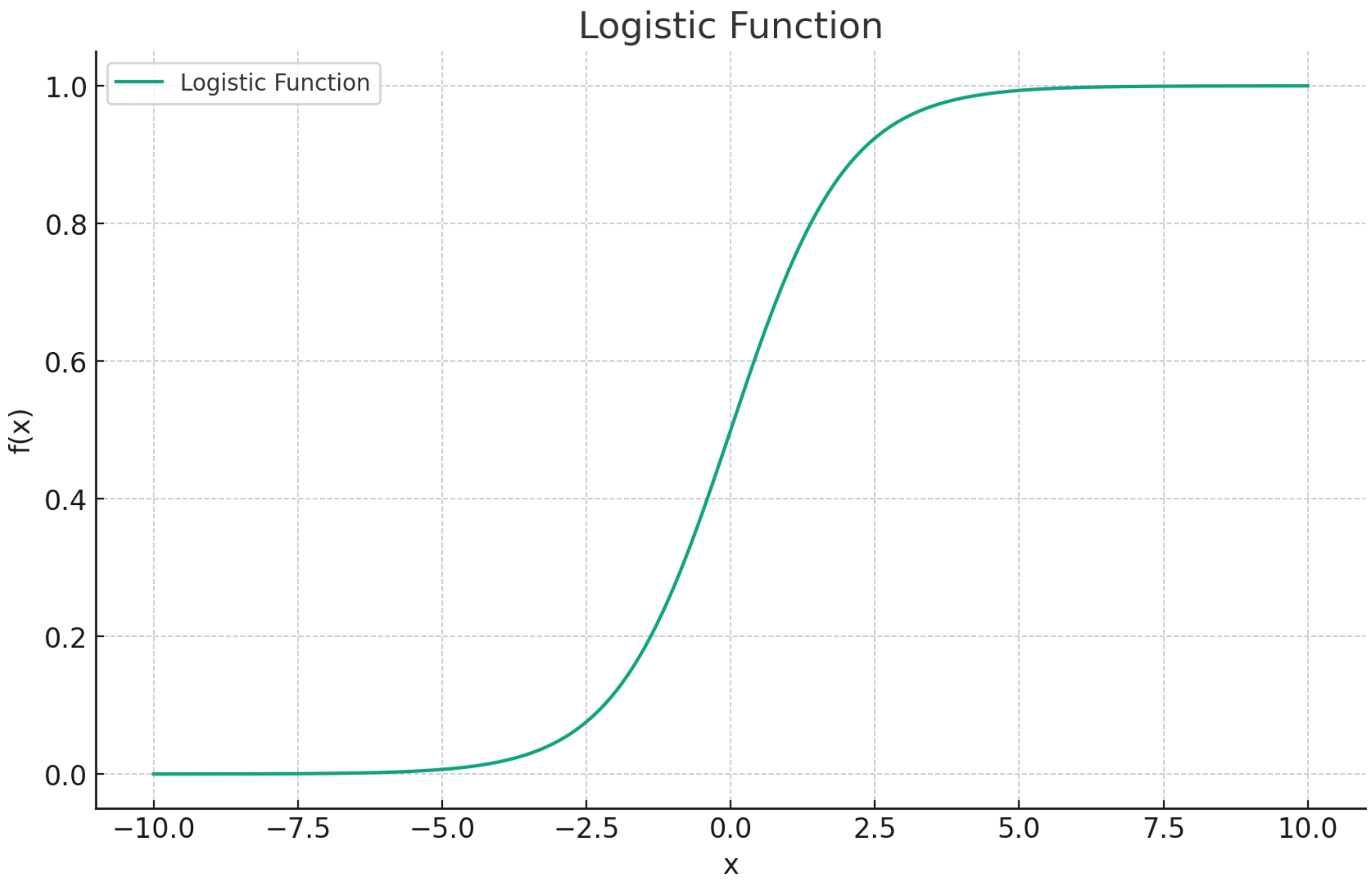
  - NO crucial: simplemente será necesario utilizar errores estándar robustos a la presencia de heterocedasticidad.
- En un modelo lineal, los valores de $E\ (Y|X = x)$ (que son probabilidades) NO están restringidos a estar comprendidos entre cero y uno.
  - se pueden tener probabilidades mayores que uno o menores que cero
  - el cambio en la probabilidad esperada de $Y = 1$ puede no
  - tener sentido

Por qué no usar el modelo de regresión lineal



Regressão Logística

- La función Logística, $f(x) = \frac{e^{(\mathrm{x})}}{1+e^{(\mathrm{x})}}$, asume valores entre 0 y 1, para cualquier x entre -∞ y +∞



$$P(y=1|x) = \frac{e^{(\mathrm{x})}}{1+e^{(\mathrm{x})}}$$

$$P(y=0|x) = 1 - \frac{e^{(\mathrm{x})}}{1+e^{(\mathrm{x})}}$$

$$P(y=0|x) = \frac{1}{1+e^{(\mathrm{x})}}$$

Modelo de regresión lógistica

- Función de Bernoulli: La función de Bernoulli describe la probabilidad de una variable aleatoria binaria.

Si Y es una variable binaria, entonces $P(Y = y) = \pi^{y}(1-\pi)^{1-y}$, donde π es la probabilidad de $(Y = 1)$.

$$E(Y) = p$$
$$Var(Y) = \pi(1-\pi)$$

- En la regresión logística, modelamos p la probabilidad de que $Y = 1$ como

$$\pi = e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)} / 1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}$$

- La función de verosimilitud L para un conjunto de datos con observaciones independientes es el producto de las probabilidades individuales:

$$L(\beta) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

donde p_i es la probabilidad modelada para la i-ésima observación.

$$\ell(\beta) = \sum_{i=1}^{n} [y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)]$$

MODELO LOGIT - ESTIMACIÓN POR MÁXIMA VEROSIMILITUD

$$Y_i = \begin{cases} 1 & si \quad características \\ 0 & no \quad características \end{cases}$$

$$P(Y_i = y_i) = \pi^{y_i}(1-\pi)^{1-y_i}$$

$$L = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$\ln L = \sum_{i=1}^{n} y_i \ln \pi_i + \sum_{i=1}^{n}(1-y_i)\ln(1-\pi_i) = \sum_{i=1}^{n} y_i \ln\frac{\pi_i}{(1-\pi_i)} + \sum_{i=1}^{n}\ln(1-\pi_i)$$

$$= \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) - \sum_{i=1}^{n}\ln\left(1 + e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}\right)$$

$$\frac{\partial \ln L}{\partial \beta_0} = 0 \Rightarrow \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}}{1 + e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}}$$

$$\frac{\partial \ln L}{\partial \beta_1} = 0 \Rightarrow \sum_{i=1}^{n} X_{1i}Y_i = \sum_{i=1}^{n} \frac{x_{1i}e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}}{1 + e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}}$$

.

.

$$\frac{\partial \ln L}{\partial \beta_k} = 0 \Rightarrow \sum_{i=1}^{n} X_{ki}Y_i = \sum_{i=1}^{n} \frac{x_{ki}e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}}{1 + e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki}}}$$

$\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$

se hallan resolviendo este sistema de (k+1) ecuaciones NO LINEALES por métodos numéricos (Newton Raphson, Scoring,...)

Regressão Logística

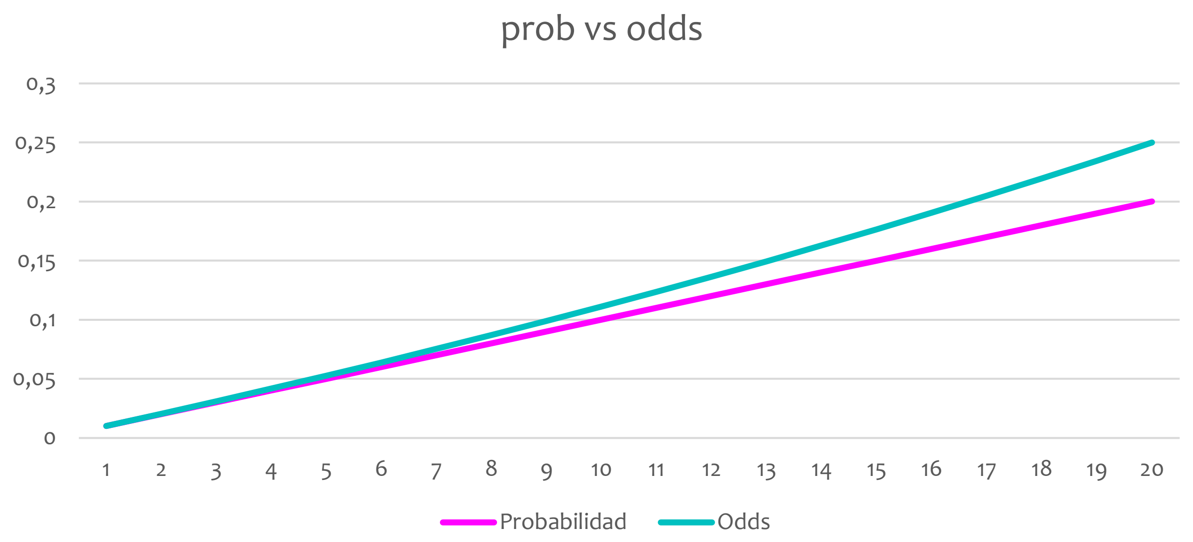
- No presume la existência de homogeneidad de varianza y normalidade de los resíduos
- $Z = ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$
- Donde π indica la probabilidad de ocurrencia de un determinado evento de interés, X representa el vector de variables explicativas (o independientes) y α y β son los parámetros del modelo

Regresión logística

- El término $ln\left(\frac{\pi}{1-\pi}\right)$ se denomina logit y el término $(p/1\ -\ p)$ representa la chance (odds) de ocurrencia del evento de interés
  - Ejemplo:
  - Si p = 0,50, la chance de ocurrencia del evento será 1 (1 a 1)
  - Si p = 0,75, la chance de ocurrencia del evento será 3 (3 a 1)

PROBABILIDAD Y ODDS

Probabilidad	Odds
0,01	0,01
0,02	0,02
0,03	0,031
0,04	0,042
0,05	0,053
0,06	0,064
0,07	0,075
0,08	0,087
0,09	0,099
0,1	0,111
0,11	0,124
0,12	0,136
0,13	0,149
0,14	0,163
0,15	0,176
0,16	0,19
0,17	0,205
0,18	0,22
0,19	0,235
0,2	0,25



Para valores muy pequeños, (menores a 0.1), no hay casi diferencias entre probabilidad y odds, pero para valores más grandes esto no es así.

Probabilidad de Y=1

$$\pi_Y = \frac{\#\ casos\ con\ (Y=1)}{total\ de\ casos}$$

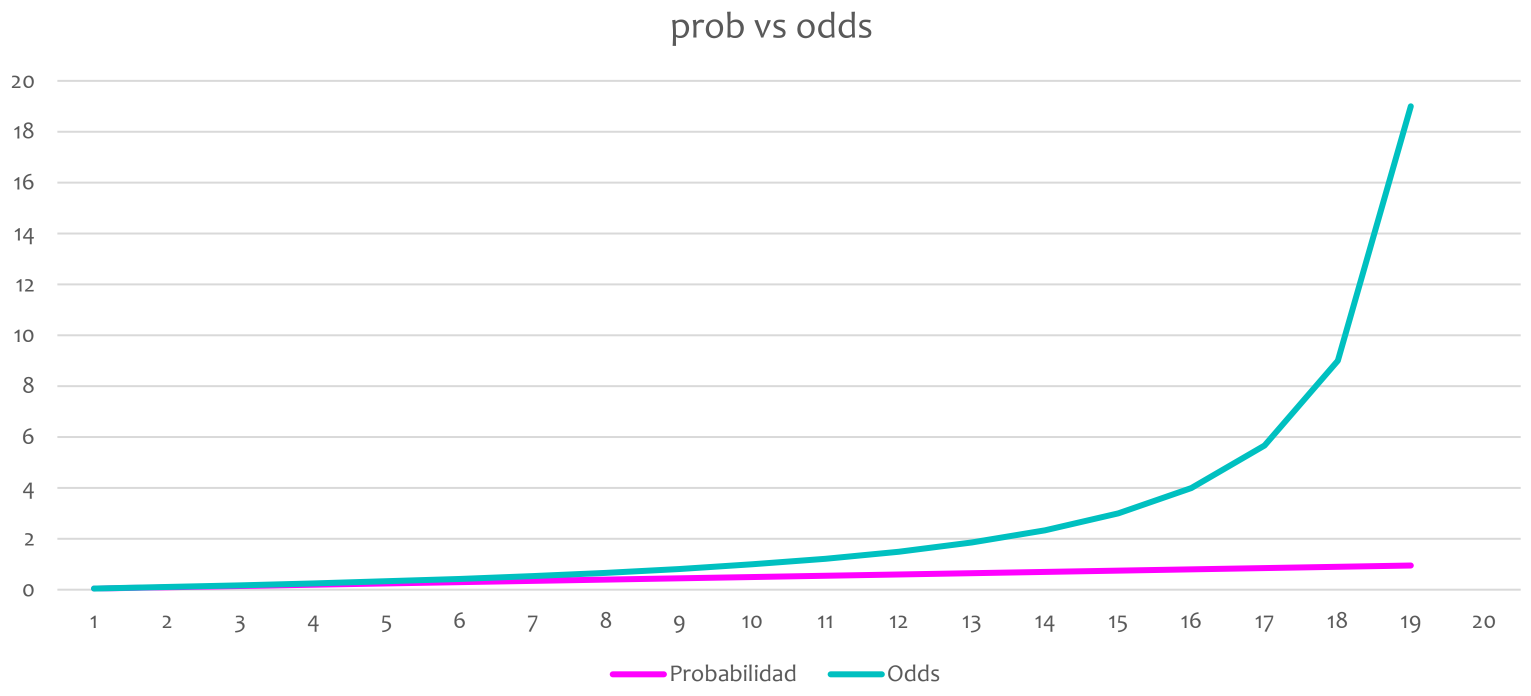
$$0 < \pi_Y < 1$$

ODDS a favor de Y=1

$$odds_Y = \frac{\pi_Y}{1-\pi_Y} = \frac{\#\ casos\ con\ (Y=1)}{\#\ casos\ con\ (Y=0)}$$

$$0 < odds_Y < \infty$$

Probabilidad	Odds
0,05	0,052631579
0,1	0,111111111
0,15	0,176470588
0,2	0,25
0,25	0,333333333
0,3	0,428571429
0,35	0,538461538
0,4	0,666666667
0,45	0,818181818
0,5	1
0,55	1,222222222
0,6	1,5
0,65	1,857142857
0,7	2,333333333
0,75	3
0,8	4
0,85	5,666666667
0,9	9
0,95	19



- **Para probabilidades mayores a 0.2 se aprecia que la diferencia entre probabilidad y odds se acentúa**
- **A probabilidades mayores a 0.5 les corresponden odds mayores a 1. Una probabilidad de 0.8 tiene asociado un odds de 4.**

Regresión Logística – Ejemplo

- Probabilidad de que un cliente compre una suscripción a una revista por correo directo

  - $prob(event) = \frac{1}{1+e^{-(1{,}143+0{,}452X_1+0{,}029X_2-0{,}242X_3)}}$

- Dónde
  - X_1 = sexo (1 para mujeres y 0 para hombres);
  - X_2 = edad;
  - X_3 = estado civil (1 para soltero y 0 para casado).

Regressão Logística – Exemplo

- Supongamos una persona de sexo femenino, de 40 años y casada, la probabilidad es
  - $prob(event) = \frac{1}{1+e^{-(1{,}143+0{,}452\cdot 1+0{,}029\cdot 40-0{,}242\cdot 0)}} = 0{,}47$
- Si fuera hombre, la probabilidad sería
  - $prob(event) = \frac{1}{1+e^{-(1{,}143+0{,}452\cdot 0+0{,}029\cdot 40-0{,}242\cdot 0)}} = 0{,}02$
- La razón de riesgo, en función del sexo, viene dada por
  - $\widehat{RR} = \frac{0{,}47}{0{,}02} = 27{,}59$
- Esto significa que la mujer tendría casi 28 veces más probabilidades de comprar la suscripción a la revista que el hombre
  - Sólo es aplicable cuando se pueden especificar todas las variables independientes y cuando el análisis se centra en cada observación

INTERPRETACIÓN

La interpretación de los $\boldsymbol{\beta_j}$ no es la pendiente como en el caso de un modelo de regresión lineal.

β_j representa el cambio en el logaritmo natural de los odds al incrementar X_j en una unidad dejando el resto de las variables constantes.

PROB

$$\pi = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}$$

LOGIT

$$log\, \frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

β_j es el cambio en el logaritmo de los odds al incrementar X_j en una unidad dejando el resto de variables constantes.

ODDS

$$\frac{\pi}{1-\pi} = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}$$

OR asociado a X_j

$$\frac{odds\,(X_j+1)}{odds\,(X_j)} = e^{\beta_j}$$

e^{β_j} es el OR asociado a X_j dejando al resto de variables constantes

Interpretación de los Odds, Odds ratio, π en el contexto de la regresión logística

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = g(x),$$

$$g(x) = \beta_0 + \beta_1 x$$

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)},$$

$$\begin{aligned} \ln\left(\frac{\pi(x+1)}{1-\pi(x+1)}\right) - \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) &= \beta_0 + \beta_1(x+1) - (\beta_0 + \beta_1 x) \\ &= \beta_1. \end{aligned}$$

$$\exp(\beta_1) = \frac{\dfrac{\pi(x+1)}{1-\pi(x+1)}}{\dfrac{\pi(x)}{1-\pi(x)}}.$$

$$\frac{\pi(x+1)}{1-\pi(x+1)} = \exp(\beta_1)\frac{\pi(x)}{1-\pi(x)}$$

```r
# Estimando un modelo de regresión logistica ----
model <- glm(deserta ~ promInicial + proced + carrera + Genero,
             data = data,
             family = binomial(link = "logit"))

summary(model)
```

```
Call:
glm(formula = deserta ~ promInicial + proced + carrera + Genero,
    family = binomial(link = "logit"), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2079  -0.5110  -0.2342  -0.0489   3.7367

Coefficients:
                  Estimate Std. Error z value             Pr(>|z|)
(Intercept)       14.94141    0.40336  37.042 < 0.0000000000000002 ***
promInicial       -4.90310    0.11092 -44.204 < 0.0000000000000002 ***
procedOTROS        0.21195    0.06803   3.116              0.00184 **
carreraEconomia    0.43399    0.17257   2.515              0.01191 *
carreraIngenieria  1.08816    0.13993   7.776  0.00000000000000747 ***
Genero1            0.53405    0.06913   7.726  0.00000000000001111 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10864.4  on 9915  degrees of freedom
Residual deviance:  6545.4  on 9910  degrees of freedom
AIC: 6557.4

Number of Fisher Scoring iterations: 6
```

```
Call:
glm(formula = deserta ~ promInicial + proced + carrera + Genero,
    family = binomial(link = "logit"), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2079  -0.5110  -0.2342  -0.0489   3.7367

Coefficients:
                  Estimate Std. Error z value             Pr(>|z|)
(Intercept)       14.94141    0.40336  37.042 < 0.0000000000000002 ***
promInicial       -4.90310    0.11092 -44.204 < 0.0000000000000002 ***
procedOTROS        0.21195    0.06803   3.116              0.00184 **
carreraEconomia    0.43399    0.17257   2.515              0.01191 *
carreraIngenieria  1.08816    0.13993   7.776  0.00000000000000747 ***
Genero1            0.53405    0.06913   7.726  0.00000000000001111 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10864.4  on 9915  degrees of freedom
Residual deviance:  6545.4  on 9910  degrees of freedom
AIC: 6557.4

Number of Fisher Scoring iterations: 6
```

$$\ln\left(\frac{\pi(x+1)}{1-\pi(x+1)}\right) - \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1(x+1) - (\beta_0 + \beta_1 x)$$
$$= \beta_1.$$

```
> # Interpretación de los coeficientes ----
>
> ## log-odds: Especificación del modelo como log(π/1-π) = \beta_0 + \beta_1*x_1 + ....
> coef(model)["promInicial"]
promInicial
  -4.903097
> # Un aumento de una unidad en el promedio inicial reduce el logaritmo de los
> # odds de desertar en 4.9 con todo lo demás constante
```

```
Call:
glm(formula = deserta ~ promInicial + proced + carrera + Genero,
    family = binomial(link = "logit"), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2079  -0.5110  -0.2342  -0.0489   3.7367

Coefficients:
                  Estimate Std. Error z value             Pr(>|z|)
(Intercept)       14.94141    0.40336  37.042 < 0.0000000000000002 ***
promInicial       -4.90310    0.11092 -44.204 < 0.0000000000000002 ***
procedOTROS        0.21195    0.06803   3.116              0.00184 **
carreraEconomia    0.43399    0.17257   2.515              0.01191 *
carreraIngenieria  1.08816    0.13993   7.776  0.00000000000000747 ***
Genero1            0.53405    0.06913   7.726  0.00000000000001111 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10864.4  on 9915  degrees of freedom
Residual deviance:  6545.4  on 9910  degrees of freedom
AIC: 6557.4

Number of Fisher Scoring iterations: 6
```

$$\frac{\pi(x+1)}{1-\pi(x+1)} = \exp(\beta_1)\frac{\pi(x)}{1-\pi(x)}$$

```
> log_odds <- predict(model, data[1,], type = "link")
> odds <- exp(log_odds)
> odds*coef(model)["promInicial"]
        1
-5.38908
> # Un aumento de una unidad en el promedio inicial reduce los odds de desertar en
> # 5.3 con todo lo demás constante
> # Para la observacion 2
> log_odds <- predict(model, data[2,], type = "link")
> odds <- exp(log_odds)
> odds*coef(model)["promInicial"]
         1
-0.4643235
> # Un aumento de una unidad en el promedio inicial reduce los odds de desertar en
> # 0.46 con todo lo demás constante
```

ODDS RATIO

$$\pi = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \qquad odds = \frac{\pi}{1-\pi} = e^{\beta_0+\beta_1 X}$$

- El OR compara los odds de dos grupos A y B o dos situaciones A y B.
- $odds_y(A) = \frac{\pi_y(A)}{1-\pi_y(A)}$
- $OR_{A,B} = \frac{odds_y(A)}{odds_y(B)} = \frac{\frac{\pi_y(A)}{1-\pi_y(A)}}{\frac{\pi_y(B)}{1-\pi_y(B)}}$
- Si X es dicotómica (X=1 si característica, X=0 si no)
  - $OR_X = \frac{odds_y(X=1)}{odds_y(X=0)} = \frac{e^{\beta_0}e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$
- Si X es continua
  - $OR_X = \frac{odds_y(X+1)}{odds_y(X)} \frac{e^{\beta_0}e^{\beta_1(x+1)}}{e^{\beta_0}e^{\beta_1(x)}} = e^{\beta_1}$

ODDS RATIO

$$OR_X = \frac{odds_y(X=1)}{odds_y(X=0)} = \frac{e^{\beta_0}e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

ACCIDENTES DE TRÁFICO FATALES

- ***Odds Ratio*** **lluvia**

Lluvia	n	Fatal	No fatal	odds	OR
Sin lluvia	340,657	4,348	336,309	0.0129	1.00
Con lluvia	22,780	356	22,424	0.0159	1.23
Total	363,437	4,704	358,733	0.0131	

- ***Odds Ratio*** **Area**

Area	n	Fatal	No fatal	odds	OR
Urbano	335,951	2,573	333,378	0.0077	1.00
Rural	27,486	2,131	25,355	0.0840	10.89
Total	363,437	4,704	358,733	0.0131	

$OR_{lluvia} > 1$

- La lluvia es un factor de riesgo.
- Los odds de que un accidente sea fatal es un 23% más alto con lluvia que sin lluvia.

El odds de un accidente con muertos en área rural es

- 10.89 veces el odds del área urbana!

ODDS Y OR

Si tenemos más de dos categorías

Por ejemplo Ciudades: Bogotá, Cali, Medellín, Barranquilla, Bucaramanga

Se escoge una base y se hallan los cocientes con la base en el denominador.

Odds Ratio ciudades

Ciudad	n	Fatal	Fatal (%)	OR	(95% IC)
Medellín	75,918	338	0.45%	1.00	
Cali	49,128	249	0.51%	1.14	(0.97-1.34)
Bogotá	67,299	612	0.91%	2.05	(1.80-2.34)
Barranquilla	8,741	81	0.93%	2.09	(1.64-2.67)
Bucaramanga	6,383	94	1.47%	3.34	(2.66-4.21)

Note que a pesar de que Barranquilla y Bucaramanga son las que menos accidentes fatales tienen, resultan ser las ciudades con más alto riesgo para los accidentes con víctimas mortales.

El odds de accidentes fatales en Bogotá es 2.05 veces el odds de accidentes fatales en Medellín, y el odds de accidentes fatales en Bucaramanga es 3.34 veces el odds de accidentes fatales en Medellín.

ODDS RATIO

```
> # Para la observación 1
> p0 <- predict(model, data[1,] %>% mutate(promInicial = promInicial), type = "response")
> odds_0 <- (p0/(1-p0))
> p1 <- predict(model, data[1,] %>% mutate(promInicial = 1 + promInicial), type = "response")
> odds_1 <- (p1/(1-p1))
> odds_1/odds_0
          1
0.007423558
> exp(coef(model)["promInicial"])
promInicial
0.007423558
>
> # Por cualquier método el odds de desertar se reduce en 99.25%
> # cuando aumenta el promedio inicial en una unidad
>
> # Para la observación 2
> p0 <- predict(model, data[2,] %>% mutate(promInicial = promInicial), type = "response")
> odds_0 <- (p0/(1-p0))
> p1 <- predict(model, data[2,] %>% mutate(promInicial = 1 + promInicial), type = "response")
> odds_1 <- (p1/(1-p1))
>
> odds_1/odds_0
          1
0.007423558
> 
```

ODDS RATIO

- Note que OR_X es constante y no depende de X a diferencia de π y odds que sí dependen de X.
- Los resultados anteriores se mantienen si el modelo tiene más variables explicativas, siempre y cuando éstas no dependan de X.
- Si hay interacción entre X_1 y X_2
- $odds = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 X_1 * X_2}$
- $OR_{X_1} = e^{\beta_1} \times e^{\beta_3 X_2}$
- OR_{X_1} no es constante y depende ahora de los valores de X_2

Efectos marginales: Usando derivadas parciales en el modelo con escala probabilística

$$\pi = \frac{e^{\beta_0+\beta_1 x_1}}{1+e^{\beta_0+\beta_1 x_1}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1)}}$$

$$\frac{\partial y}{\partial x_1} = \frac{\beta_1}{(1+e^{-(\beta_0+\beta_1 x_1)})^2} \cdot \frac{1}{e^{\beta_0+\beta_1 x_1}}$$

$$\frac{\partial y}{\partial x_1} = \beta_1 \pi^2 \cdot \frac{1}{\pi/(1-\pi)}$$

$$\frac{\partial y}{\partial x_1} = \beta_1 \pi^2 \cdot \frac{(1-\pi)}{\pi}$$

$$\frac{\partial y}{\partial x_1} = \beta_1 \pi (1-\pi)$$

EFECTO MARGINAL

$$\pi = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}} = \frac{e^{\beta' X}}{1+e^{\beta' X}}$$

El efecto marginal de X_j es la derivada con respecto a X_j de la función.

$$\frac{\partial \pi(Y)}{\partial X_j} = \frac{d\,\pi(\beta' X)}{d(\beta' X)} * \frac{d(\beta' X)}{d X_j} = \frac{e^{\beta' X}}{\left(1+e^{\beta' X}\right)^2} * \beta_j$$

- El efecto marginal de X_j en el modelo logit depende de los valores de X_j a diferencia del modelo de regresión lineal donde es constante.

- Para interpretar el modelo , se recomienda calcular efectos marginales en los valores medios de las variables o en algunos valores pertinentes.

```
Call:
glm(formula = deserta ~ promInicial + proced + carrera + Genero,
    family = binomial(link = "logit"), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2079  -0.5110  -0.2342  -0.0489   3.7367

Coefficients:
                  Estimate Std. Error z value            Pr(>|z|)
(Intercept)       14.94141    0.40336  37.042 < 0.0000000000000002 ***
promInicial       -4.90310    0.11092 -44.204 < 0.0000000000000002 ***
procedOTROS        0.21195    0.06803   3.116              0.00184 **
carreraEconomia    0.43399    0.17257   2.515              0.01191 *
carreraIngenieria  1.08816    0.13993   7.776  0.00000000000000747 ***
Genero1            0.53405    0.06913   7.726  0.00000000000001111 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10864.4  on 9915  degrees of freedom
Residual deviance:  6545.4  on 9910  degrees of freedom
AIC: 6557.4

Number of Fisher Scoring iterations: 6
```

$$\frac{\partial y}{\partial x_1} = \beta_1 \pi (1 - \pi)$$

```
> ## Probabilidad: especificación del modelo como π = exp(\beta_0 + \beta_1*x_1 + ....) / 
 ....))
> # Para la observacion 1
> prob <- predict(model, data[1,], type = "response")
> prob*(1-prob)*coef(model)["promInicial"]
        1
-1.223041
> # Un aumento de una unidad en el promedio inicial reduce la probabilidad de desertar en
> # 12200 puntos básicos
>
> # Para la observacion 2
> prob <- predict(model, data[2,], type = "response")
> prob*(1-prob)*coef(model)["promInicial"]
         1
-0.3874631
> # Un aumento de una unidad en el promedio inicial reduce la probabilidad de desertar en
> # 3874 puntos básicos
```

Efectos marginales: Por qué son necesarios ?

- En el modelo lineal simple , como por ejemplo $y = \beta_0 + \beta_1$age+ $\beta_2$male, podemos interpretar fácilmente los coeficientes.

- Recordemos que los coeficientes representan los efectos marginales (derivadas parciales), esto es, la razón de cambio unitario de una variable independiente y la variable dependiente.

- Cuando el modelo contiene expresiones no lineales, por ejemplo,

$$y = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 male$$

$$\frac{\partial E[y|age, male]}{\partial age} = \beta_1 + 2\beta_2 age$$

- No hay un efecto único de la edad, el efecto depende de la edad; es decir, un efecto a los 20 años, otro a los 50, etc. Basta con introducir los números de la edad en la expresión anterior para obtener el efecto a diferentes edades.

Efectos marginales: Por qué son necesarios ?

- Podemos escribir el modelo logístico como $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 age + \beta_2 male$
  - Los parámetros estimados están en la escala log-odds, que, aparte del signo, no tienen ninguna interpretación útil.
- En la ecuación anterior, β_1 es el efecto de la edad en las probabilidades logarítmicas del resultado, no en la probabilidad, que suele ser lo que nos interesa.
- Con el modelo logístico podemos presentar los odd ratios $\left(e^{\beta_1}\ y\ e^{\beta_2}\right)$, pero los odd ratios son frecuentemente mal interpretados como relacionados a riesgos y probabilidades.

Ejemplo sin covariantes. Suponga que la probabilidad de muerte en un grupo de control es 0.40. La probabilidad de muerte en el grupo de tratamiento es 0.20

- El odd ratio es $\frac{{}^{0,2}/_{0,8}}{{}^{0,4}/_{0,6}} = 0.375$. El tratamiento reduce la chance de muerte por un factor de 0.375. O inversamente, las chances de muerte son 2.67 mayores en el grupo de control $\left({}^{1}/_{0.375}\right)$
- Pero esto no es un riesgo relativo, el riesgo relativo es $\frac{0.2}{0.4} = 0.5$. La probabilidad de muerte es reducido a la mitad en el grupo de tratamiento.

- Con las odds ratios y los riesgos relativos, no tenemos una idea de la magnitud.
  - El mismo ejemplo, pero ahora la probabilidad de muerte en el grupo de control es de 0.0004 y de 0.0002 en el grupo de tratamiento. EL odds ratio sigue siendo de 0,375 y el riesgo relativo sigue siendo de 0,5. Las magnitudes son, por supuesto, muy diferentes.
- Un periodista podría seguir diciendo que, por ejemplo, comer brotes de brócoli a diario reduce la probabilidad de morir de cáncer a la mitad. ¡¡¡A la mitad!!!Pero si se enterara de que la reducción es de (0.0004-0.0002) 0.0002 o 0,02 puntos porcentuales, probablemente no correrá a comer brócoli todos los días.
- Por otro lado, una diferencia de 20 puntos porcentuales parece bastante impresionante.
- Los efectos marginales son una forma de presentar los resultados como diferencias en las probabilidades, lo cual es más informativo que los odds ratios y los riesgos relativos.

BONDAD Y AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICO

¿Que tan bueno es nuestro modelo ?

- El ajuste de los modelos logísticos se evalúa en generan en términos compartidos con otros modelos con más/menos predictores.

- Estas medidas de comparación se basan en la log verosimilitud (log-likelihood) del modelo, que es una magnitud que se obtiene dado el procedimiento de estimación en regresión logística.

- Entre las medidas de ajuste usualmente se consideran:
  - Criterio de máxima verosimilitud
  - Devianza (deviance)
  - Test de razón de verosimilitud (likelihood ratio test)
  - R^2
  - Criterios de información

Criterio de máxima verosimilitud

- Al igual que en regresión múltiple, podemos usar estos valores observados y predichos para evaluar el ajuste del modelo. La medida que usamos es log-likelihood (logaritmo de la razón de verosimilitud):

$$log - likelihood = \sum_{i=1}^{n} \left[y_i log\pi(x_i) + (1 - y_i) log\big(1 - \pi(x_i)\big) \right]$$

- Se basa por tanto en la suma de las probabilidades asociadas con los resultados estimados y los valores reales. El estadístico log-likelihood es análogo a la suma de cuadrados residual en la regresión múltiple en el sentido de que es un indicador cuánta información sin explicar queda en la variable respuesta tras haber ajustado el modelo.

Grandes valores del log-likelihood indican un pobre ajuste del modelo, cuanto mayor sea este valor, más variabilidad sin explicar queda en el modelo.

Devianza

- La desviación de un modelo ajustado compara la log-verosimilitud del modelo ajustado con la log-verosimilitud de un modelo con n parámetros que se ajusta perfectamente a las n observaciones.

- **Puede demostrarse que la verosimilitud de este modelo saturado es igual a 1, lo que produce una log-verosimilitud igual a 0**. Por lo tanto, la desviación del modelo de regresión logística es:

$$D = -2\sum_{i=1}^{n}\left[Y_i \ln\left(\widehat{\pi}_i\right) + \left(1 - Y_i\right)\ln\left(1 - \widehat{\pi}_i\right)\right] = -2\ln\frac{L_M}{L_S}$$

- D sigue una distribución chi cuadrado.
- Los valores más pequeños indican un mejor ajuste a media que el modelo ajustado (L_M) se desvía menos del modelo saturado (L_S) .
- Entienda por saturado un modelo con ajuste perfecto.
- Cuando se evalúa sobre una distribución chi cuadrado, los valores no significativos indican muy poca variación inexplicable y, por tanto, un buen ajuste del modelo. Por el contrario, un valor significativo de chi-cuadrado indica que una cantidad significativa de la varianza no tiene explicación.
- Cuando el modelo saturado no está disponible (un caso común), la desviación se calcula simplemente como $-2 \cdot Ln\ L_M$, y la referencia a la probabilidad de registro del modelo saturado se puede eliminar de todo lo que sigue sin daño.

PRUEBA DEL LIKELIHOOD RATIO

- Las pruebas de hipótesis están basadas en la verosimilitud y en las propiedades asintóticas de los estimadores de máxima verosimilitud.
- Prueba del likelihood Ratio:
  - Compara un modelo Full con un modelo restringido a través de sus verosimilitudes.

$$Ho: Modelo\ Reducido\ (R)\ vs\ Ha: Modelo\ Full(F)$$

$$LR = 2\left(ln\left(L_F\right) - ln\left(L_R\right)\right) \sim \chi^2_{(p-q)}$$

- Prueba para la significancia Global del modelo

$$Ho: Modelo\ No\ significativo\ vs\ Ha: Modelo\ significativo$$
$$Ho: Modelo\ con\ constante\ vs\ Ha: Modelo\ con\ todas\ las\ variables$$

$$LR = 2\left(ln\left(L_M\right) - ln\left(L_0\right)\right) \sim \chi^2_{(p-1)}$$

PSEUDO R-CUADRADO

- Versiones análogas al R-cuadrado del modelo de regresión lineal se han propuesto en la literatura.
- Los **PSEUDO R-CUADRADO** se basan en la relación entre el valor del verosimilitud del modelo L_M y el valor de verosimilitud de un modelo sin variables, es decir, del modelo únicamente con la constante, L_0
- En la literatura estadística los pseudo R-cuadrados no son muy utilizados.
- No hay evidencia convincente de que modelos con mayores pseudo R-cuadrados tengan mejores resultados.

- McFadden

$$R^2_{McF} = 1 - \frac{\ln(L_M)}{\ln(L_0)}$$

- Cox&Snell

$$R^2_{ML} = 1 - \left(\frac{L_0}{L_M}\right)^{2/n}$$

- Nagelkerke

$$R^2_N = \frac{1 - \left(\frac{L_0}{L_M}\right)^{2/n}}{1 - (L_M)^{2/n}}$$

- El pseudo R^2 en porcentajes
- El pseudo R^2 no es realmente interpretable, pero puede utilizarse para comparar modelos.

Model	$p+1$	R^2_{McF}	$R^2_{\text{McF,adj}}$	$R^2_{\text{Cox-Snell}}$	$R^2_{\text{Nagelkerke}}$
0:null	1	0	0	0	0
1:cars	2	4.8 %	4.6 %	5.0 %	7.6 %
2:cars+wind	3	6.6 %	6.2 %	6.8 %	10.3 %
red:tempdiff	3	6.9 %	6.5 %	7.1 %	10.8 %
3:cars+zerodiff	3	11.7 %	• 11.3 %	11.6 %	17.9 %
4:cars*wind	4	8.2 %	7.1 %	7.7 %	12.9 %
5:cars+tempdiff	4	11.8 %	• 11.3 %	11.9 %	18.1 %
6:cars*wind+zerodiff	5	13.2 %	12.4 %	13.2 %	20.0 %
oslo:cars*wind+tempdiff	6	13.7 %	● 12.7 %	13.7 %	20.8 %

CRITERIOS DE INFORMACIÓN

- Uno de los criterios para seleccionar, se basa en la perdida de información. Los criterios de información más populares son el AIC y el BIC.
- Estos tienen en cuenta tanto la verosimilitud como el número de parámetros p del modelo.

  - $AIC = -2logL_M + 2p$
  - $BIC = -2logL_M + log(n) \times p$

Model	df	AIC	BIC
0:null	1	535.8	543.1
1:cars	2	515.2	523.7
2:cars+wind	3	507.6	520.3
red:tempdiff	3	505.8	518.4
3:cars+zerodiff	3	480.3	**492.9**
4:cars*wind	4	500.6	517.5
5:cars+tempdiff	4	481.4	498.3
6:cars*wind + zerodiff	5	476.2	497.3
oslo:cars*wind+tempdiff	6	**475.4**	500.7

Bondad en la predicción

En ocasiones queremos usar el modelo para clasificar futuros objetos como "suceso" o "fracaso", dependiendo de las probabilidades dadas por sus valores de x.

Lo más fácil es clasificar utilizando:

	Predicted ($\hat{Y}_i$)		
True (Y_i)	Failure ($\hat{p}_i \leq 0,5$)	Success ($\hat{p}_i > 0.5$)	Total
Failure ($Y_i = 0$)	true negative	false positive	TN + FP
Success ($Y_i = 1$)	false negative	true positive	FN + TP
Total	TN + FN	FP + TP	n

Nota: hay situaciones en las que podríamos querer un umbral diferente. Queremos examinar la proporción de las observaciones que están correctamente . Esto puede hacerse utilizando la matriz de confusión:

- **Sensitividad** (Sensitivity): es la proporción de los verdaderos éxitos que han sido correctamente clasificados como suceso (tasa de verdaderos positivos)

$$\text{Sensitivity} = \frac{TP}{FN + TP}$$

- **Especificidad** (Specificity): Es la proporción de verdaderos fracasos que han sido correctamente clasificados como fracasos (tasa de verdaderos negativos)

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- **Exactitud** (Accuracy): es la proporción global que se ha clasificado correctamente

$$\text{Accuracy} = \frac{TP + TN}{n}$$

- **Precisión** (Precision): es la proporción de los éxitos predichos que son verdaderos éxitos

$$\text{Precision} = \frac{TP}{FP + TP}$$

The largest model: `I(cars/1000)*windspeed + tempdiff`:

True	Predicted 0	Predicted 1	total	Correctly classified	
0	374	12	386	96.9 %	specificity
1	90	24	114	21.1 %	sensitivity
total	464	36	500		

Accuracy: $(374 + 24)/500 = 79.6\,\%$. Precision: $24/36 = 66.7\,\%$.

The best (BIC) model: `I(cars/1000) + zerodiff`:

True	Predicted 0	Predicted 1	total	Correctly classified	
0	369	17	386	95.6 %	specificity
1	92	22	114	19.3 %	sensitivity
total	461	39	500		

Accuracy: $(369 + 22)/500 = 78.2\,\%$. Precision: $22/39 = 56.4\,\%$

- Sensitivity= $\frac{TP}{FN+TP}$
- Specificity= $\frac{TN}{TN+FP}$
- Accuracy= $\frac{TP+TN}{n}$
- Precision= $\frac{TP}{FP+TP}$
- F1–score=2 $\cdot \frac{Sentivity \cdot Precision}{Sentivity+Precision}$

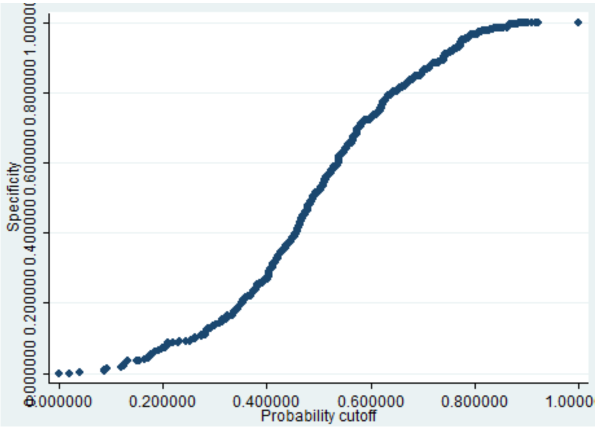
True	Predicted 0	Predicted 1
0	TN	FP
1	FN	TP

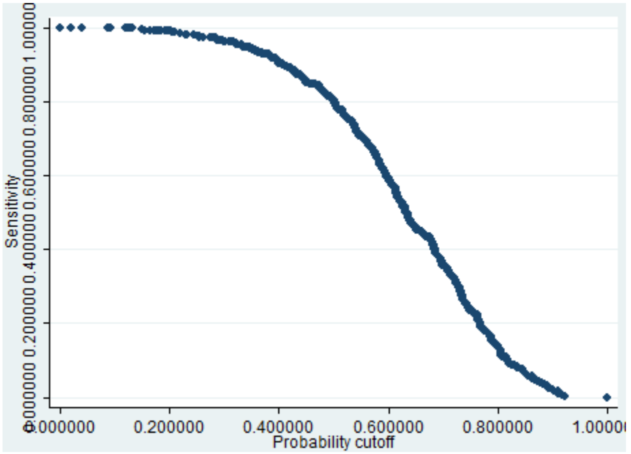
Otros puntos de corte pueden utilizarse. La pertinencia de éstos cortes puede evaluarse si se tienen en cuenta por ejemplo los costos sociales o económicos de una clasificación incorrecta.

```
. estat class, cutoff (0.7)

Logistic model for trabaja

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       153            46  |        199
     -     |       275           279  |        554
-----------+--------------------------+-----------
   Total   |       428           325  |        753

Classified + if predicted Pr(D) >= .7
True D defined as trabaja != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   35.75%
Specificity                     Pr( -|~D)   85.85%
Positive predictive value       Pr( D| +)   76.88%
Negative predictive value       Pr(~D| -)   50.36%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   14.15%
False - rate for true D         Pr( -| D)   64.25%
False + rate for classified +   Pr(~D| +)   23.12%
False - rate for classified -   Pr( D| -)   49.64%
--------------------------------------------------
Correctly classified                        57.37%
--------------------------------------------------
```

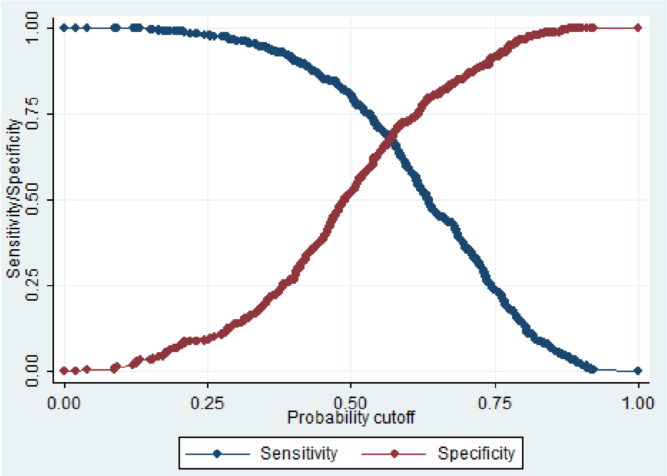
```
. estat class, cutoff (0.6)

Logistic model for trabaja

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       253            88  |        341
     -     |       175           237  |        412
-----------+--------------------------+-----------
   Total   |       428           325  |        753

Classified + if predicted Pr(D) >= .6
True D defined as trabaja != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   59.11%
Specificity                     Pr( -|~D)   72.92%
Positive predictive value       Pr( D| +)   74.19%
Negative predictive value       Pr(~D| -)   57.52%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   27.08%
False - rate for true D         Pr( -| D)   40.89%
False + rate for classified +   Pr(~D| +)   25.81%
False - rate for classified -   Pr( D| -)   42.48%
--------------------------------------------------
Correctly classified                        65.07%
--------------------------------------------------
```

Especificidad versus punto de corte de la probabilidad





Aumentar el punto de corte, incrementa el porcentaje de negativos verdaderos y disminuye el porcentaje de positivos verdaderos.



$$Especificidad = \frac{negativos\ correctos}{total\ negativos}$$

$$Sensitividad = \frac{positivos\ correctos}{total\ positivos}$$

Advertencias

- No hay una manera fácil de "castigar" la adición de más variables x.

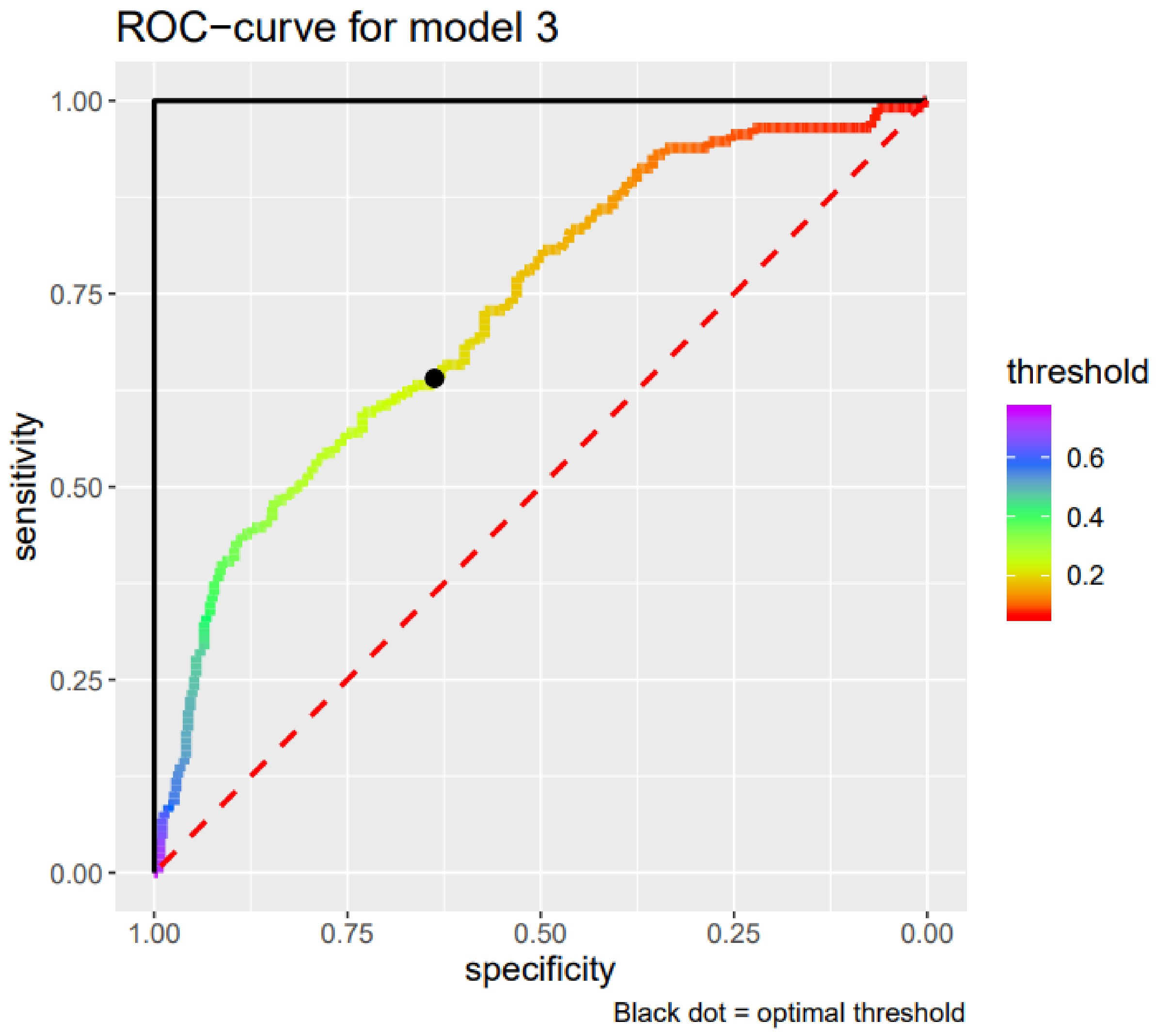
- Los modelos más grandes suelen tener valores más altos al predecir los mismos datos que se utilizaron al estimar el modelo, debido al sobreajuste.

- Si el objetivo principal del modelo es clasificar las observaciones futuras, deberíamos validar en un conjunto de datos separado, que no intervenga en el ajuste del modelo.

- Por otra parte, si el modelo ni siquiera puede predecir sus propios datos, no es un modelo muy bueno.

Curva ROC

- Cambiando el valor del umbral de 0,5 por otro valor podemos cambiar la especificidad y la sensitividad. Podemos hacer que cualquiera de ellas sea tan grande como queramos, pero a costa de que la otra sea pequeña.
- A menudo puede ser importante tener tanto una gran especificidad como una gran sensitividad.
- Podríamos probar todos los valores de umbral posibles y calcular la especificidad y la sensibilidad para cada uno de ellos.
- Elegir el valor del umbral que haga que tanto la especificidad como la sensibilidad sean lo más grandes posible, al mismo tiempo.
- Esta capacidad del modelo para separar dos categorías es ilustrada en la curva ROC *(Receiver Operating Characteristics)*



- Obsérvese la escala invertida en el eje x. (gráfico 1-especificidad)
- El negro sólido es la curva ROC ideal. La roja discontinua es la del modelo nulo.
- Utilizando el umbral $\hat{\pi}_i > 0{,}2215$ se obtiene una sensibilidad del 64,0 % y una especificidad del 63,7%.

Área bajo la curva (AUC)

- El área bajo la curva ROC (AUC) mide lo cerca que estamos de la curva ideal. Área ideal = 1.

- Modelo nulo (lanzar una moneda) = 0,5. Las áreas inferiores a 0,5 son peores que el modelo "lanzar una moneda".

- El AUC es la probabilidad de que, si se toma un par de observaciones al azar, en las que una es un éxito y la otra un fracaso, el éxito tenga una mayor probabilidad predicha de ser un éxito que el fracaso. Por tanto, el AUC da la probabilidad de que el modelo clasifique correctamente esos pares de observaciones.

- El modelo más grande tiene el mayor AUC, como se esperaba.
- Sólo el modelo 6 no es significativamente diferente de "Oslo".
- Por otra parte, las diferencias entre los modelos 3, 5, 6 y Oslo son pequeñas.

Model	AUC	95 % C.I.	P-value
0:null	50.0	(50.0, 50.0)	< 0.001
1:cars	64.8	(59,3, 70.3)	< 0.001
2:cars+wind	68.6	(63.5, 73.7)	0.002
red:tempdiff	63.8	(58.8, 68.9)	< 0.001
3:cars+zerodiff	72.8	(67.5, 78.1)	0.03
4:cars*wind	69.5	(64.3, 74,6)	0.003
5:cars+tempdiff	73.0	(67.7, 78.2)	0.046
6:cars*wind + zerodiff	74.9	(69.9, 79.8)	0.12
oslo:cars*wind+tempdiff	75.5	(70.5, 80.4)	-

PRUEBA DE WALD

- Prueba de Wald para significancias individuales

- Se basa en el hecho que un estimador de máxima verosimilitud $\hat{\theta}_{MV}$ es asintóticamente normal y cuya matriz de varianza se determina a través de las segundas derivadas parciales de $Ln(L)$

En el caso del modelo logit, $\frac{\partial^2 \ln L}{\partial \beta_j \, \beta_k} = -\sum_i^n x_{ji} x_{ki} \; \pi_i (1 - \pi_i)$

$$\hat{\beta}_{MV} \sim AN\left(\beta, Var(\hat{\beta})\right)$$

$$\widehat{Var}(\hat{\beta}) = \left(X'\hat{V}X\right)^{-1}; \quad \hat{V} = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \\ & & \ddots \; \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

$$H_o: \beta_j = 0 \quad vs. \quad H_a: \beta_j \neq 0$$

$$z = \frac{\hat{\beta}_j}{e.e\,(\hat{\beta}_j)} \sim N(0,1)$$

Bajo Ho

Rechace si $|z| > K$

PREDICCIÓN DE PROBABILIDADES

A partir de los estimadores del modelo logit, podemos predecir la probabilidad para valores particulares de las variables explicativas

ESTIMADOR PUNTUAL para $\pi_{Xp} = P\left(Y = 1/X_1 = x_{1p}, X_2 = x_{2p}, \dots, X_k = x_{kp}\right)$

$$\hat{\pi}_{xp} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1p} + \cdots + \hat{\beta}_k X_{kp}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1p} + \cdots + \hat{\beta}_k X_{kp}}}$$

El INTERVALO DE CONFIANZA se halla en dos pasos

PASO 1 I. de C. para $\beta' X_p$ $\qquad L \leq \beta' X_p \leq U$

PASO 2 I. de C. para $\frac{e^{\beta' x_p}}{1+e^{\beta' x_p}}$ $\qquad \frac{e^L}{1+e^L} \leq \pi \leq \frac{e^U}{1+e^U}$

$$(L, U) = \hat{\beta}' X_p \pm K \sqrt{Var(\hat{\beta}' X_p)} \; ; Var(\hat{\beta}' X_p) = (1, X_{1p}, \dots, X_{kp}) \underbrace{\widehat{Var}(\hat{\beta})}_{(X'\hat{V}X)^{-1}} (1, X_{1p}, \dots, X_{kp})'$$

$$\boxed{\text{I. de C. para } \pi_{Xp} : \left(\frac{e^L}{1+e^L}, \frac{e^U}{1+e^U}\right)}$$

PREDICCIÓN DE PROBABILIDADES

A partir de los estimadores del modelo logit, podemos predecir la probabilidad para valores particulares de las variables explicativas

trabaja	Coef.	Std. Err.	z	P>|z|
numhijos	-1.429763	.1912293	-7.48	0.000
edad	-.0567239	.0112771	-5.03	0.000
mbach	1.081808	.1981083	5.46	0.000
ing	-.0314051	.0076496	-4.11	0.000
_cons	3.386345	.528165	6.41	0.000

Por ejemplo, la probabilidad de una mujer de 40 años que tiene 1 hijo, no tiene bachillerato y tiene un ingreso de 20, trabaje y aporte a los ingresos del hogar es

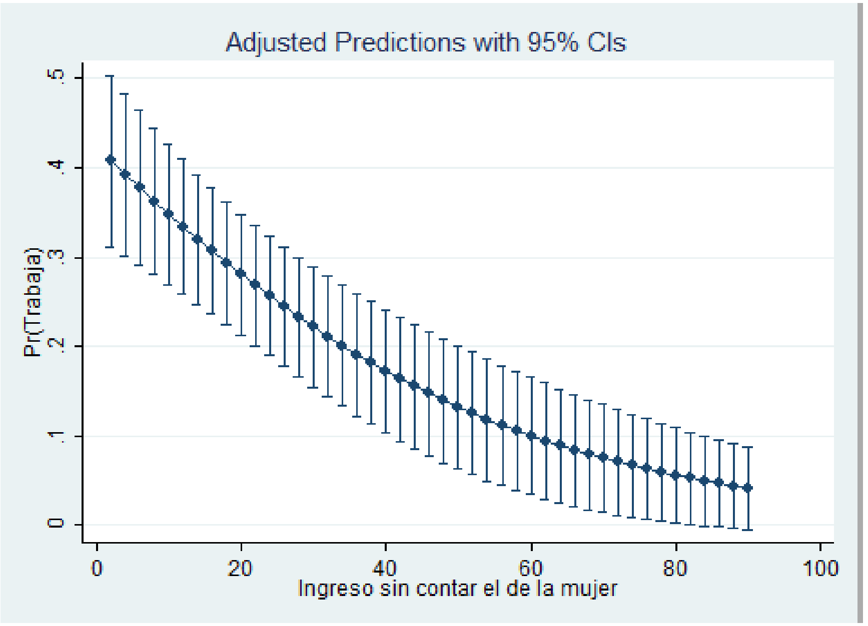
$$\hat{\pi} = \frac{e^{\hat{\beta}_0+1\hat{\beta}_1+40\hat{\beta}_2+20\hat{\beta}_4}}{1+e^{\hat{\beta}_0+1\hat{\beta}_1+40\hat{\beta}_2+20\hat{\beta}_4}} = 0.281$$

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	.2808045	.03453	8.13	0.000	.213127	.348482

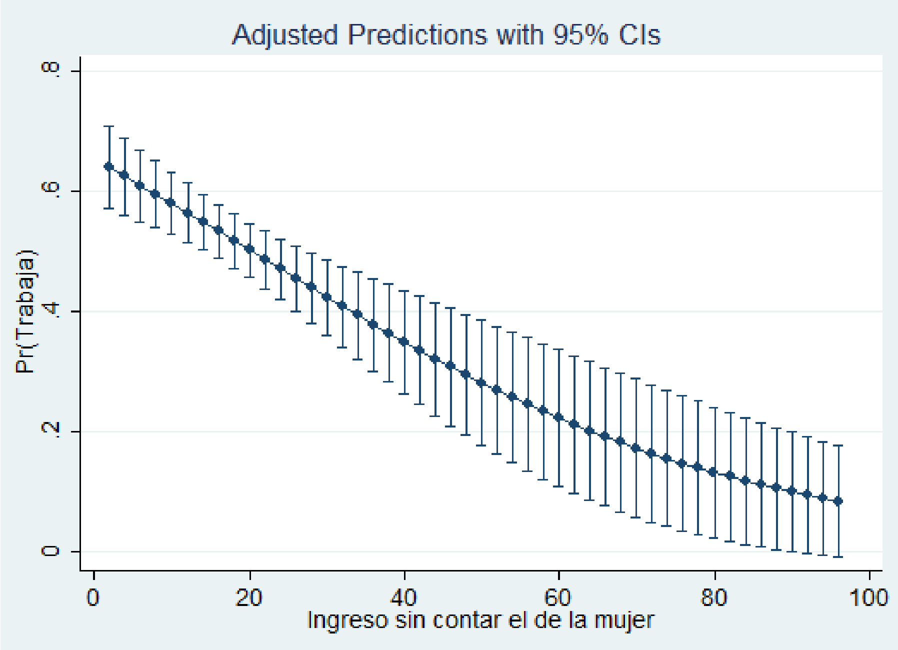
Gráficas de Probabilidades

Probabilidades en función del Ingreso

Las probabilidades y los intervalos de confianza varían de acuerdo a los valores particulares escogidos.

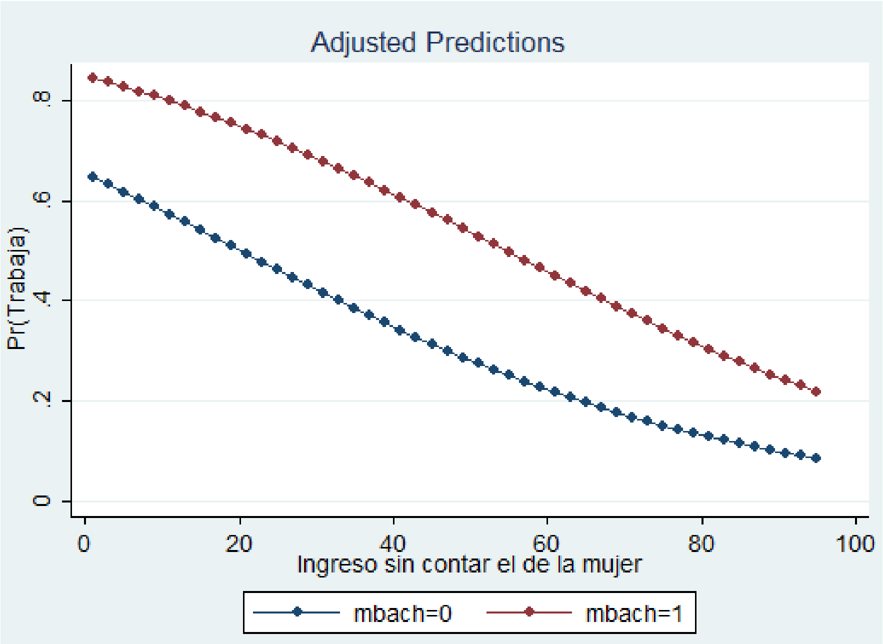


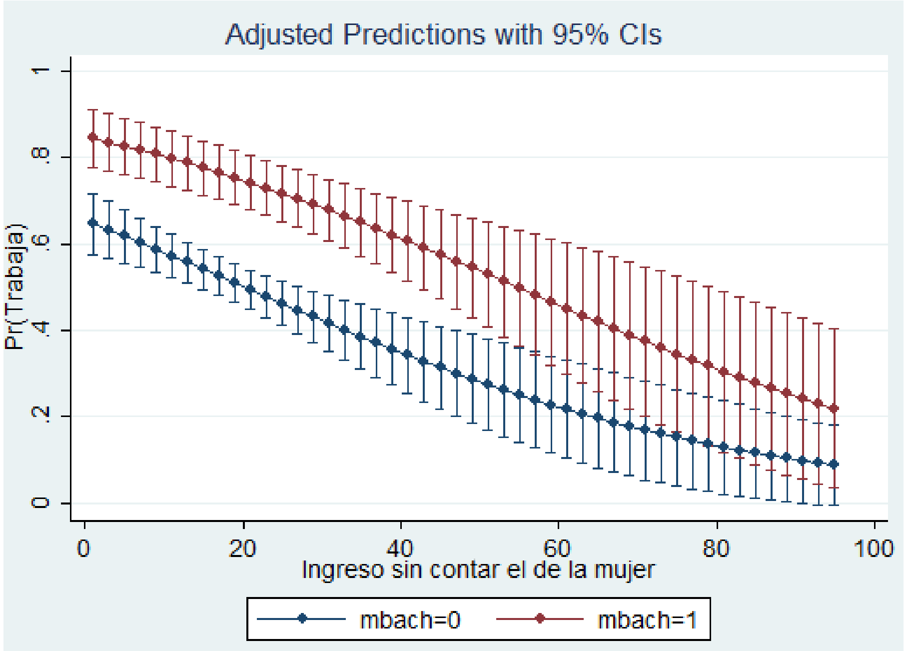
edad=40, 1 hijo, sin bachillerato



sin bachillerato, el resto en los valores medios

Efecto de tener o no bachillerato - Efecto de una variable dummy.





DIAGNOSTICO

Resumen de los supuestos de regresión logística

- **Todos los residuos deben ser independientes (no correlacionados entre sí).** Las observaciones deben ser independientes entre sí. Los datos no deben tener observaciones duplicadas, medidas repetidas u otros resultados correlacionados.

- **Linealidad:** los logits (los registros naturales de las probabilidades de que la variable dependiente sea igual a 1) deben estar linealmente relacionados con las variables numéricas independientes continuas. Esto se puede probar observando gráficos residuales y trazando logits contra cada variable independiente.

- **Multicolinealidad:** No debería haber multicolinealidad en nuestro modelo. Esto se define y prueba exactamente de la misma manera que en la regresión lineal múltiple.

- **Valores atípicos y observaciones influyentes:** no queremos que nuestra muestra incluya observaciones que puedan ser valores atípicos que influyan en los resultados de nuestro modelo.

Residuos de Pearson

Normalización simple, ya que Y_i sigue una distribución de Bernoulli con $E(Y_i) = \pi_i$ y $\text{Var}(Y_i) = \pi_i(1 - \pi_i)$

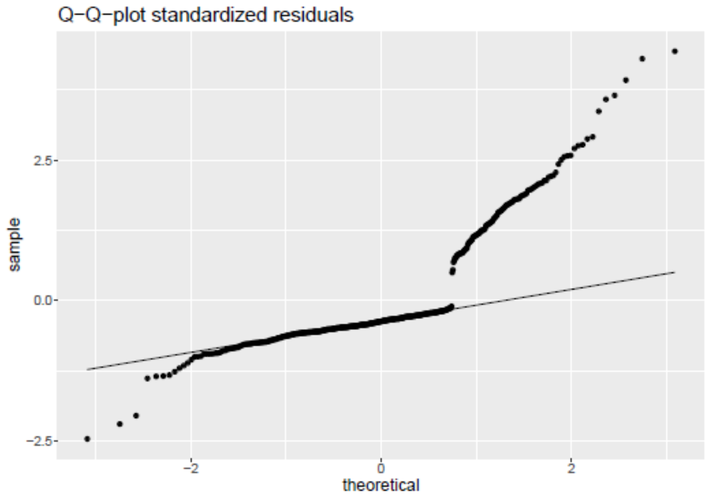
$$r_i = \frac{Y_i - \pi_i}{\sqrt{\pi_i\left(1-\pi_i\right)}} \rightarrow r_i^2 \sim \chi^2$$

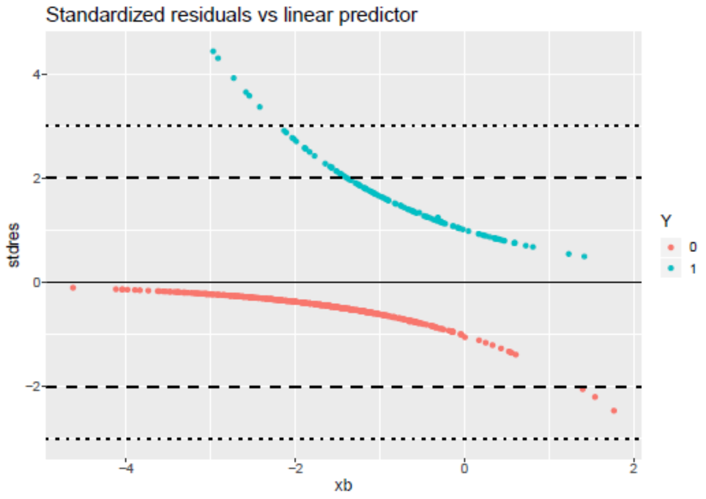
Residuos estudentizados

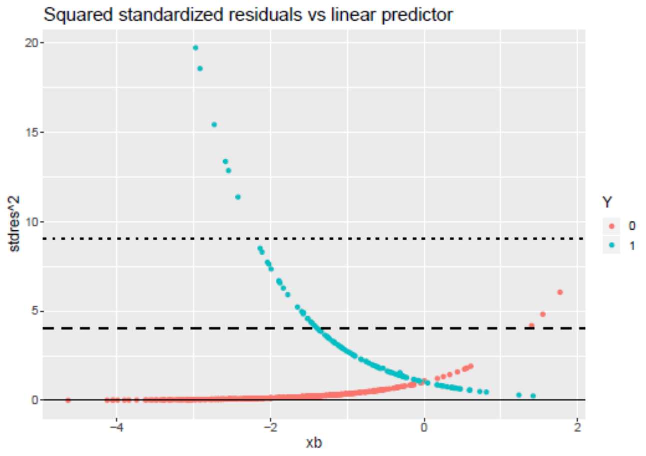
Al igual que en la regresión lineal, podemos estudentizar los residuos utilizando el apalancamiento:

$$r_i = \frac{Y_i - \pi_i}{\sqrt{\pi_i\left(1-\pi_i\right)\left(1-v_{ii}\right)}} \rightarrow r_i \sim N\left(0,1\right), n \rightarrow grande$$

Si $|r_i| > \left|Z_{\alpha/2}\right| \approx 2$ o > 3 podría considerarse sospechosamente grande. Los gráficos de $r_i\ vs\ X\beta$ pueden ser útiles, aunque a veces es más revelador trazar sus cuadrados, por ejemplo, $r_i^2 vs\ X\beta$







- No se distribuye normalmente.
- Residuos más grandes cuando Yi = 1.
- Tendencias extrañas (¡una característica, no un error!)

Residuos de desviación

Desde que $0 \leq L(\hat{\beta}) \leq 1$, la devianza $D = -2Ln\, L(\hat{\beta})$ es una suma de valores no negativos y podemos reescribirla como una suma de cuadrados:

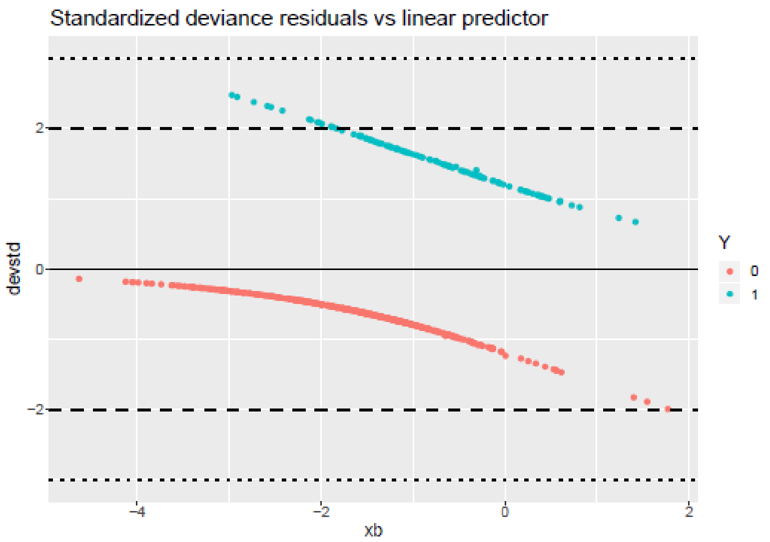
$$D = -2\sum_{i=1}^{n}\left(Y_i \ln \pi_i + (1-Y_i)\ln(1-\pi_i)\right)$$

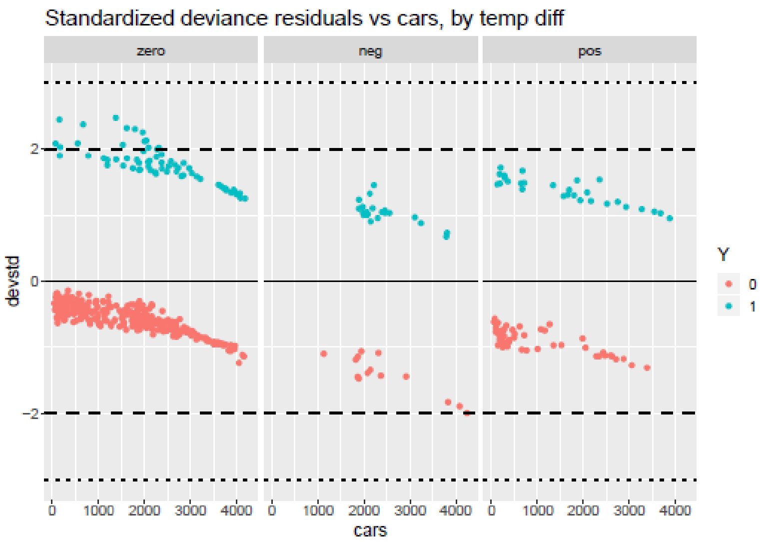
$$D = \sum_{i|Y_i=1} 2\ln\frac{1}{\pi_i} + \sum_{i|Y_i=0} 2\ln\frac{1}{1-\pi_i} = \sum_{i=1}^{n} d_i^2$$

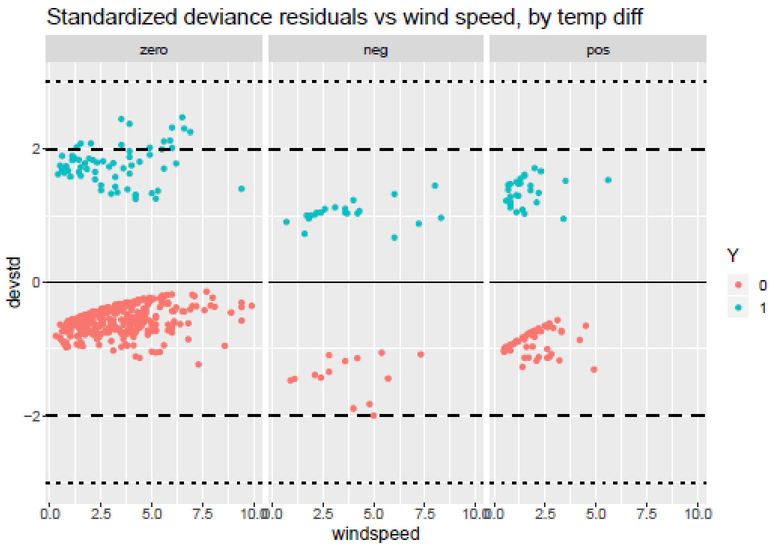
$$d_i = \begin{cases} -\sqrt{2\ln\dfrac{1}{1-\pi_i}}, Y_i = 0 \\ +\sqrt{2\ln\dfrac{1}{\pi_i}}, Y_i = 1 \end{cases}$$

El residuo de desviación será pequeño si $Y_i = 0$ y π_i es cercano a cero, o si $Y_i = 1$ y π_i es cercano a uno. En caso contrario, será grande.

Los residuos de desviación pueden estandarizarse como $d_i = {}^{1}/_{\sqrt{1-v_{ii}}}$.
Si el valor absoluto es > 2 o > 3 puede considerarse demasiado grande.







- **Se comporta mucho mejor.**
- **Algunos ligeros problemas para predecir las altas concentraciones en la diferencia de temperatura cero, pero no son alarmantes.**

Observaciones influyentes

Podemos medir la influencia de las observaciones individuales en las estimaciones β de forma similar a la regresión lineal.

Distancia de Cook: Existe una versión de la distancia de Cook para la regresión logística:

$$D_i^{Cook} = \frac{r_i^2}{p+1} \cdot \frac{v_{ii}}{1 - v_{ii}}$$

donde ri son los residuos estudentizados. Podríamos considerar casos influyentes aquellos con DCook > 1 o > 4/n.

- Algunas observaciones con gran influencia en las estimaciones.
- Algunas de ellas eran observaciones con un gran apalancamiento.
- Pero las diferencias de temperatura negativas son un problema, especialmente cuando la concentración de PM10 no es alta.