

Trabajo Práctico 5 - Confusores y modificadores de efecto

María Eugenia Fontecha

26 octubre, 2020

La base de datos *death.csv* corresponde a un estudio donde los casos fueron recién nacidos que fallecieron por injurias. Los controles eran recién nacidos de edad similar. Las variables de la base de datos son las siguientes:

- byear: año de nacimiento
- injflag: 1: muerte, 0: vivo
- pnclate: 1: cuidado perinatal en las primeras 48hs del nacimiento, 0: sin cuidado en las primeras 48 hs
- YoungW: 1: madres adolescentes, 0: madres mayores de edad.

```
setwd(path)
death = read.csv('death.csv')
summary(death)
```

```
##      X      byear      injflag      pnclate
## Min.   : 1    Min.   :1986    Length:5624    Length:5624
## 1st Qu.:1407  1st Qu.:1989    Class :character    Class :character
## Median :2812  Median :1991    Mode  :character    Mode  :character
## Mean   :2812  Mean   :1991
## 3rd Qu.:4218  3rd Qu.:1993
## Max.   :5624  Max.   :1995
##      YoungMo
## Length:5624
## Class :character
## Mode  :character
##
##
##
```

Reemplazo con 0 y 1 los datos para que queden como en el enunciado y paso a factor las variables dicotómicas.

```
injflag <- death$injflag
death$injflag[injflag == 'No'] <- 0
death$injflag[injflag == 'Yes'] <- 1
death$injflag <- as.factor(death$injflag)
str(death$injflag)
```

```
## Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
pnclate <- death$pnclate
death$pnclate[pnclate == 'No'] <- 0
death$pnclate[pnclate == 'Yes'] <- 1
death$pnclate <- as.factor(death$pnclate)
str(death$pnclate)
```

```
## Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
YoungMo <- death$YoungMo
death$YoungMo[YoungMo == 'No'] <- 0
death$YoungMo[YoungMo == 'Yes'] <- 1
death$YoungMo <- as.factor(death$YoungMo)
str(death$YoungMo)
```

```
## Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

1. Obtenga el OR sin ajustar (crudo) y luego el OR de Mantel-Haenszel de madres adolescentes vs. madres mayores para muerte, ajustado por año de nacimiento, ya que las tecnologías de cuidado neonatal mejoraron con los años, y en años más recientes podría haber habido una reducción de la mortalidad.

Genero la tabla de contingencia de madres adolescentes.

```
tabla_youngmo <- table(death$YoungMo, death$injflag)
tabla_youngmo = tabla_youngmo[rev(order(rownames(tabla_youngmo))) ,
                             rev(order(colnames(tabla_youngmo)))]
kable(tabla_youngmo, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	1	0
1	258	1610
0	249	3507

Realizo el test de Chi Cuadrado para ver si hay una asociación entre el factor de exposición madre adolescente y el evento muerte del recién nacido.

```
#Test de Chi Cuadrado, con corrección de continuidad (porque la tabla es de 2x2)
chi_youngmo <- chisq.test(tabla_youngmo)
chi_youngmo
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tabla_youngmo
## X-squared = 77.585, df = 1, p-value < 2.2e-16
```

Hay una asociación estadísticamente significativa entre la variable madre adolescente y la muerte del recién nacido, según el test de Chi Cuadrado, ya que el p-valor es menor a 0.05.

Calculo el OR crudo.

```
OR_YoungMo <- odds.ratio(tabla_youngmo, level = 0.95)
kable(OR_YoungMo, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.256658	1.869641	2.724224	0

```
#p-value
print(OR_YoungMo$p)
```

```
## [1] 8.03409e-18
```

A partir del OR crudo, se concluiría que el ser madre adolescente es un factor de riesgo para la muerte del recién nacido por injurias, aumentando el riesgo del evento 2.26 veces, con un intervalo de confianza del 95% entre 1.87 y 2.72.

Agrupo por año de nacimiento.

```
t_ym_year <- as.data.frame(table(death$YoungMo, death$byear, death$injflag))
colnames(t_ym_year)=c("young_mo", "year", "injflag", "f")
kable(t_ym_year, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

young_mo	year	injflag	f
0	1986	0	306
1	1986	0	112
0	1987	0	275
1	1987	0	109
0	1988	0	261
1	1988	0	109
0	1989	0	432
1	1989	0	178
0	1990	0	419
1	1990	0	190
0	1991	0	495
1	1991	0	217
0	1992	0	429
1	1992	0	222
0	1993	0	408
1	1993	0	213
0	1994	0	338
1	1994	0	184
0	1995	0	144
1	1995	0	76
0	1986	1	28
1	1986	1	14
0	1987	1	28
1	1987	1	13
0	1988	1	16
1	1988	1	22
0	1989	1	30
1	1989	1	27
0	1990	1	29
1	1990	1	35
0	1991	1	35
1	1991	1	38
0	1992	1	28
1	1992	1	36
0	1993	1	26
1	1993	1	33
0	1994	1	19
1	1994	1	30
0	1995	1	10
1	1995	1	10

Genero la tabla de contingencia para cada año.

```
t_estrato<-(xtabs(f~young_mo+injflag+year, data=t_ym_year))
```

Calculo el OR por estrato.

```
odds <- apply(t_estrato,3,odds.ratio)
```

1986

```
kable(odds[['1986']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	1.365093	0.6397399	2.796167	0.3678041

1987

```
kable(odds[['1987']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	1.170896	0.5358924	2.439348	0.7167554

1988

```
kable(odds[['1988']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	3.281503	1.577984	6.964823	0.0007765

1989

```
kable(odds[['1989']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.181467	1.20979	3.918966	0.0064957

1990

```
kable(odds[['1990']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.657266	1.528338	4.651843	0.0002531

1991

```
kable(odds[['1991']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.473478	1.477212	4.153068	0.0003341

1992

```
kable(odds[['1992']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.481172	1.430753	4.341952	0.0006017

1993

```
kable(odds[['1993']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.427813	1.368225	4.348097	0.0016024

1994

```
kable(odds[['1994']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.894705	1.52872	5.606223	0.0005597

1995

```
kable(odds[['1995']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	1.889454	0.6730877	5.307663	0.2225904

Análisis de Homogeneidad

Verifico si los OR entre los diferentes estratos son constantes. Las hipótesis del test de homogeneidad son:

- H_0 : los OR son homogéneos
- H_1 : los OR no son homogéneos.

Si no rechazo la hipótesis nula, entonces el año de nacimiento podría ser un factor confusor y luego tengo que hacer el cálculo del OR ajustado, calcular la diferencia porcentual entre este y el OR global crudo y realizar el test de asociación. Si rechazo la hipótesis nula, entonces el año de nacimiento es un modificador de efecto y tengo que reportar el OR por estrato.

```
BreslowDayTest(x = t_estrato)
```

```
##
## Breslow-Day test on Homogeneity of Odds Ratios
##
## data: t_estrato
## X-squared = 8.3956, df = 9, p-value = 0.4948
```

El test de homogeneidad dio un p-valor mayor a 0.05, por lo que no se rechaza la hipótesis nula. No hay diferencia estadísticamente significativa entre los OR de los distintos estratos.

Odds ratio global y test de Mantel-Haenszel

Calculo el OR global ajustado y realizo el test para ver si hay una asociación significativa entre el ser madre adolescente y la muerte por injurias del recién nacido, cuando se ajusta por año de nacimiento. El test de Mantel-Haenszel verifica si ajustando por el factor confusor, sigue habiendo una asociación significativa entre el factor de exposición y el evento. Las hipótesis son:

- H_0 : $OR = 1$
- H_1 : $OR \neq 1$.

```
mantel <- mantelhaen.test(t_estrato)
mantel
```

```
##
## Mantel-Haenszel chi-squared test with continuity correction
##
```

```
## data: t_estrato
## Mantel-Haenszel X-squared = 78.433, df = 1, p-value < 2.2e-16
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
## 1.895243 2.742317
## sample estimates:
## common odds ratio
## 2.279771
```

Calculo la diferencia porcentual entre el OR ajustado y el OR crudo. Veo si hay una diferencia mayor al 10% entre el OR crudo y el OR ajustado por año de nacimiento.

```
(mantel$estimate - OR_YoungMo$OR)/OR_YoungMo$OR
```

```
## common odds ratio
## 0.01024232
```

2. ¿Es razonable obtener un único OR entre los estratos? Justifique. ¿Cómo reportaría los resultados?

Considero que es razonable obtener un OR único entre los estratos ya que realizando el test de homogeneidad se obtuvo que no hay diferencia estadísticamente significativa entre ellos, por lo que no sería necesario reportar un OR para cada uno. Supongo que los resultados que obtuve son razonables ya que, si bien pudo haber habido un avance tecnológico que haya implicado una mejora en los cuidados perinatales, en el análisis realizado no se tiene en cuenta qué grupo recibió estos cuidados y cuál no, sino que solamente se está separando en madres adolescentes y madres mayores en cada año.

El OR ajustado por año de nacimiento dio 2.279771, y al compararlo con el OR crudo, de 2.256658, se puede ver que la diferencia entre ellos es menor al 10% (cercana al 1%), por lo tanto el año de nacimiento no es un factor confusor. Por esta razón, reportaría como resultados el OR crudo y su correspondiente intervalo de confianza, como se muestra a continuación.

Hay una asociación estadísticamente significativa entre ser madre adolescente y la muerte por injurias del recién nacido, siendo la primera variable un factor de riesgo. El ser madre adolescente aumenta en un 125% las chances de que el recién nacido muera por injurias, con un intervalo de confianza al 95% entre 87% y 172%.

3. Obtenga lo mismo que se pide en 1. y 2. pero ahora ajustando por cuidado perinatal. Interprete los resultados. ¿Qué reportaría en este nuevo análisis?

Agrupo por madre adolescente o mayor y cuidado perinatal.

```
t_ym_peri <- as.data.frame(table(death$YoungMo, death$pnclate, death$injflag))
colnames(t_ym_peri)=c("young_mo", "cuidado_perinatal", "injflag", "f")
kable(t_ym_peri, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```


young_mo	cuidado_perinatal	injflag	f
0	0	0	3265
1	0	0	1244
0	1	0	242
1	1	0	366
0	0	1	219
1	0	1	184
0	1	1	30
1	1	1	74

Genero tablas de contingencia.

```
t2_estrato<-(xtabs(f~young_mo+injflag+cuidado_perinatal, data=t_ym_peri))
```

Calculo el OR por estrato.

```
odds2 <- apply(t2_estrato,3,odds.ratio)
```

Sin cuidados perinatales

```
kable(odds2[['0']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	2.204767	1.782704	2.724708	0

Con cuidados perinatales

```
kable(odds2[['1']], booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

	OR	2.5 %	97.5 %	p
Fisher's test	1.629892	1.017731	2.663389	0.0378307

Test de homogeneidad

Realizo el test de homogeneidad para ver si los OR por estrato son homogéneos.

```
BreslowDayTest(x = t2_estrato)
```

```
##
## Breslow-Day test on Homogeneity of Odds Ratios
##
## data: t2_estrato
## X-squared = 1.4123, df = 1, p-value = 0.2347
```

Con un p-valor mayor a 0.05 no se rechaza la hipótesis nula, por lo que los OR por estrato son homogéneos. Esto indica que la variable cuidados perinatales no es un modificador de efecto.

OR ajustado por cuidados perinatales

Calculo el OR global, ajustado por cuidados perinatales, y realizo el test de Mantel-Haenszel.

```
mantel <- mantelhaen.test(t2_estrato)
mantel

##
## Mantel-Haenszel chi-squared test with continuity correction
##
## data:  t2_estrato
## Mantel-Haenszel X-squared = 59.479, df = 1, p-value = 1.236e-14
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.721423 2.513817
## sample estimates:
## common odds ratio
##           2.080227
```

Diferencia porcentual entre el OR crudo y el ajustado por cuidados perinatales.

```
(OR_YoungMo$OR - mantel$estimate)/OR_YoungMo$OR

## common odds ratio
##           0.0781825
```

Como la diferencia entre el OR crudo y el OR ajustado es de casi 8%, es decir menor al 10%, la variable cuidados perinatales no es un factor confusor.

Como conclusión, al calcular el OR de madres adolescentes agrupando según cuidados perinatales obtuve que los OR por estrato eran homogéneos, por lo tanto esta variable no es un modificador de efecto y no es necesario reportar resultados por estrato. Además, al calcular el OR global, ajustado por cuidados perinatales, la diferencia entre este y el OR crudo fue menor al 10%, por lo que esta variable no es un factor confusor. Entonces me quedo con los datos crudos.

Por lo tanto, hay una asociación estadísticamente significativa entre la muerte por injurias de recién nacidos y que las madres sean adolescentes, siendo el riesgo de muerte 2.26 veces mayor que en el caso de que las madres sean mayores, con un intervalo de confianza al 95% entre 1.87 y 2.72.

Tanto para el caso del año de nacimiento como para los cuidados perinatales, no tuve en cuenta los resultados del test de Mantel-Haenszel ya que según los análisis que hice, ninguna se considera un factor confusor en la asociación entre madre adolescente y muerte del recién nacido. Si este hubiese sido el caso, igualmente habría una asociación estadísticamente significativa entre ser madre adolescente y la muerte por injurias del recién nacido ajustando por año de nacimiento ya que el test de Mantel-Haenszel dio un p-valor menor a 0.05, y también habría una asociación estadísticamente significativa entre ser madre adolescente y la muerte por injurias del recién nacido ajustando por cuidados perinatales porque el test tuvo un p-valor menor a 0.05.