

# Evidencia 2: Comparativo de Modelos

Jorge Uriel Barragan Pozos

1/7/2022

## Comparativo de modelos

Se compararan 4 modelos en los incluiremos, o no, algunas variables y se estudiaran el comportamiento de las variables que resultan irrelevantes para el modelo. Utilizaremos la siguiente base de datos:

```
summary(edu_inc)
```

```
##      faminc      he      we      kl6
## Min.   : 9072   Min.   : 4.00   Min.   : 5.00   Min.   :0.0000
## 1st Qu.: 61562  1st Qu.:11.00   1st Qu.:12.00   1st Qu.:0.0000
## Median : 83013  Median :12.00   Median :12.00   Median :0.0000
## Mean   : 91213  Mean   :12.61   Mean   :12.66   Mean   :0.1402
## 3rd Qu.:112618  3rd Qu.:16.00   3rd Qu.:14.00   3rd Qu.:0.0000
## Max.   :344146  Max.   :17.00   Max.   :17.00   Max.   :2.0000
##      xtra_x5      xtra_x6
## Min.   : 2.258   Min.   : 9.367
## 1st Qu.:10.394   1st Qu.:21.945
## Median :12.601   Median :24.914
## Mean   :12.570   Mean   :25.135
## 3rd Qu.:14.827   3rd Qu.:28.185
## Max.   :20.815   Max.   :37.675
```

## Definimos los Modelos

```
mod1 <- lm(faminc~he+we,edu_inc)

mod2 <- lm(faminc~he,edu_inc)

mod3 <- lm(faminc~he+we+kl6,edu_inc)

mod4 <- lm(faminc~he+we+kl6+xtra_x5+xtra_x6,edu_inc)
```

## Comparamos los Modelos

```
stargazer(mod1, mod2, mod3, mod4, type="text", title = "Comparación modelos", align = TRUE,
          out = "tabla1.txt", flip = TRUE, digits = 1)
```

```
##
## Comparación modelos
## =====
##                               Dependent variable:
##                               -----
##                               faminc
##                               -----
##                               (1)          (2)          (3)          (4)
## -----
## he              3,131.5***          5,155.5***          3,211.5***          3,339.8***
##                  (802.9)          (658.5)          (796.7)          (1,250.0)
##
## we              4,522.6***          4,776.9***          5,868.7**
##                  (1,066.3)          (1,061.2)          (2,278.1)
##
## kl6              -14,310.9***          -14,200.2***
##                  (5,003.9)          (5,043.7)
##
## xtra_x5              888.8
##                  (2,242.5)
##
## xtra_x6              -1,067.2
##                  (1,981.7)
##
## Constant         -5,533.6          26,191.3***          -7,755.3          -7,558.6
##                  (11,229.5)          (8,541.1)          (11,162.9)          (11,195.4)
##
## -----
## Observations          428          428          428          428
## R2                    0.2          0.1          0.2          0.2
## Adjusted R2           0.2          0.1          0.2          0.2
## Residual Std. Error  40,497.9 (df = 425)  41,297.5 (df = 426)  40,160.1 (df = 424)  40,239.9 (df = 422)
## F Statistic         40.9*** (df = 2; 425)  61.3*** (df = 1; 426)  30.4*** (df = 3; 424)  18.3*** (df = 5; 422)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Primero se puede observar que en el primer modelo, el termino constante no es significativo para el modelo y el modelo explica el 20 por ciento de los datos. En el segundo modelo, que considera dos variables, vemos que ambas y el termino constante son significativas. Sin embargo, la  $R^2$  se reduce, de forma que el modelo explica solo el 10 por ciento de los datos. Para el modelo 3, se observa que incluyendo la variable kl6 que es significativa para el modelo, vuelve a incrementar la  $R^2$  y de nuevo el termino constante no es significativo. Finalmente, para el modelo 4 que incluye las variables xtra\_x5 y xtra\_x6 vemos que la  $R^2$  es muy parecida a la de los demás modelos y además las nuevas variables no son significativas para el modelo.

## ¿Cuál es la diferencia entre las variables omitidas y las variables irrelevantes?

Se puede observar los modelos 3 y 4, el cuarto incluye las variables irrelevantes y el tercero omite estas variables, la principal diferencia que se puede observar es en el estadístico F.