

## Tarea 2

Dania Bustamante, Rosana Cardona, José Casanova

2024-04-09

```
fertility_1 <- read_excel("fertility-1.xlsx")
```

### ¿Cómo afecta la fertilidad a la oferta de trabajo?

#### Pregunta 1:

Corra una regresión de weeksm1 sobre la variable indicador morekids, utilizando MCO. ¿En promedio, las mujeres con más de dos hijos trabajan menos que las mujeres con dos hijos? ¿Cuánto menos?

#### Modelo Poblacional

$$\text{weeksm1} = \beta_0 + \beta_1 \text{morekids} + u$$

#### Modelo Estimado

$$\hat{\text{weeksm1}} = \hat{\beta}_0 + \hat{\beta}_1 \text{morekids}$$

```
rg <- lm(weeksm1 ~ morekids, fertility_1)
stargazer(rg, type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               weeksm1
## -----
## morekids                      -5.387***
##                               (0.089)
##
## Constant                      21.068***
##                               (0.055)
##
## -----
## Observations                  254,654
## R2                           0.014
## Adjusted R2                   0.014
## Residual Std. Error          21.710 (df = 254652)
## F Statistic                   3,696.022*** (df = 1; 254652)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

El intercepto es 21.068. Esto significa que, en promedio, las mujeres casadas con dos o menos hijos trabajaron alrededor de 21.07 semanas en 1979. Por otra parte, el parámetro de morekids es -5.38700, lo que significa que, en promedio, las mujeres casadas que tuvieron más de dos hijos trabajaron aproximadamente 5.39 semanas menos en el año 1979 en comparación con las mujeres que tuvieron dos o menos hijos. El análisis indica que, en promedio, tener más de dos hijos tuvo un impacto negativo en la cantidad de semanas trabajadas por las mujeres casadas en 1979.

### Pregunta 2:

Explique por qué la regresión MCO estimada en (a) no es apropiada para estimar el efecto causal de la fertilidad (morekids) en la oferta laboral (weeksm1).

La regresión estimada  $\text{weeksm1} = \hat{\beta}_0 + \hat{\beta}_1 \text{morekids}$  podría no ser apropiada para estimar el efecto causal de la fertilidad (morekids) en la oferta laboral (weeksm1) debido a:

- **Omisión de variables relevantes:** La regresión puede estar omitiendo variables importantes que afectan tanto a la fertilidad como a la oferta laboral. Una de estas puede ser el nivel educativo o el ingreso familiar que podrían tener un impacto significativo en la relación entre la fertilidad y la oferta laboral. Por otra parte, el modelo solo incluye la variable explicativa morekids, es probable que otras variables relevantes estén correlacionadas tanto con morekids como con la variable explicada weeksm1. Las variables como la edad de la madre y el género de los primeros dos hijos podrían influir en la decisión de tener más hijos y en la participación laboral de la mujer. No incluir estas variables podría generar algún sesgo en las estimaciones del efecto de morekids en la oferta laboral.
- **Endogeneidad:** Pueden existir factores no observados que influyen tanto en la fertilidad como en las semanas trabajadas. Las preferencias de las mujeres hacia la una profesión y la maternidad pueden influir en ambas variables. Por lo tanto, la variable morekids puede estar correlacionada con el término de error de la regresión. Además, es posible que la oferta laboral de una mujer influya en su decisión de tener más hijos, es decir, una mujer que trabaja más semanas puede posponer la decisión de tener más hijos, generando una **simultaneidad** entre la oferta laboral y la fertilidad.

### Pregunta 3:

El conjunto de datos contiene la variable samesex, que es igual a 1 si los primeros dos hijos son del mismo sexo (niño-niño o niña-niña) y igual a 0 en caso contrario. ¿Las parejas cuyos dos primeros hijos son del mismo sexo tienen más probabilidades de tener un tercer hijo? Explique por qué samesex es un instrumento válido para la regresión IV de weeksm1 en morekids.

La variable samesex podría ser un instrumento válido para morekids, ya que es razonable pensar que las parejas que tienen dos hijos del mismo sexo podrían tener más incentivos para tener un tercer hijo en comparación con las parejas que han tenido hijos de sexos diferentes. Podemos suponer que algunas parejas pueden desear tener un hijo del otro sexo después de tener dos hijos del mismo sexo. Por lo tanto, samesex podría ser relevante para explicar la decisión de tener un tercer hijo.

Por otra parte, para asegurarnos que samesex es un instrumento válido, deberá cumplir con dos características:

1. **Relevancia:**  $\text{Cov}(Z, X) \neq 0$ , es decir, la variable instrumental debe explicar a X.
2. **Exogeneidad:**  $\text{Cov}(Z, u) = 0$  es decir, la variable instrumental no debe correlacionar con el error.

#### Pregunta 4:

Estime la regresión IV de *weeksworked* en *morekids*, utilizando *samesex* como instrumento. ¿Qué tan grande es el efecto de la fertilidad en la oferta laboral?

Estimamos la regresión  $morekids = \pi_0 + \pi_1 samesex + v$  para saber si mi variable instrumental es significativa.

```
XMCDE <- lm(morekids ~ samesex, fertility_1)
stargazer(XMCDE,type = "text")

##
## =====
##                               Dependent variable:
##                               -----
##                               morekids
## -----
## samesex                      0.068***
##                               (0.002)
##
## Constant                     0.346***
##                               (0.001)
##
## -----
## Observations                 254,654
## R2                           0.005
## Adjusted R2                  0.005
## Residual Std. Error         0.484 (df = 254652)
## F Statistic                  1,237.219*** (df = 1; 254652)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

Realizamos el t - test para verificar si la variable *samesex* es un instrumento valido:

- Hipotesis nula:  $H_0 : samesex = 0$
- Hipotesis alterna:  $H_1 : samesex_1 \neq 0$

```
t_calculado <- (0.068-0)/(0.002)
cat(t_calculado)
```

```
## 34
```

Remplazando:

$$t = \frac{0.068-0}{0.002}$$

$$t = \frac{0.068}{0.002}$$

$$t = 34$$

Si el  $t$  calculado es mayor que 1.9634 o menor que  $-1.9634$ , se rechaza la hipotesis nula. De lo contrario, no tendríamos suficiente evidencia para rechazar la hipótesis nula.

Dado que el valor calculado de  $t = 34$  es mayor en valor absoluto que 1.96, podemos concluir que cae fuera del intervalo crítico. Por lo tanto, tenemos suficiente evidencia estadística para rechazar la hipótesis nula.

Según los resultados, el parámetro de la variable `samesex` es 0.0675, lo que indica que, si los dos primeros hijos son del mismo sexo, en promedio, la probabilidad de querer tener más hijos es de 0.0675 en comparación a las mamás que tienen hijos de distinto sexo. Observando el valor `p`, éste es muy pequeño, por lo tanto, se deduce que el parámetro es estadísticamente significativo. Se concluye que el instrumento `samesex` es válido.

Corremos la regresión con Variables Instrumentales

```
rgIV <- ivreg(weeksm1 ~ morekids | samesex, data= fertility_1)
stargazer(rgIV,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               weeksm1
## -----
## morekids                     -6.314***
##                               (1.275)
##
## Constant                     21.421***
##                               (0.487)
##
## -----
## Observations                 254,654
## R2                          0.014
## Adjusted R2                 0.014
## Residual Std. Error    21.715 (df = 254652)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

El coeficiente estimado para `morekids` es -6.314. Con base en los resultados de la regresión con Variables Instrumentales y la validez del instrumento, podemos concluir que existe evidencia de que tener más de dos hijos está asociado con una reducción de aproximadamente 6.314 semanas en la cantidad de trabajo de las mujeres en el año 1979 en comparación a las mujeres que tienen menos hijos, después de controlar por los sesgos de endogeneidad utilizando `samesex` como instrumento.

### Pregunta 5:

¿Los resultados de la pregunta 4 cambian cuando se incluyen las variables `agem1`, `black`, `hispan` y `othrace` en la regresión de la oferta laboral (tratando estas variables como exógenas)? Explique por qué o por qué no.

Las variables `agem1`, `black`, `hispan` y `othrace` son características que no deberían afectar la cantidad de trabajo de manera directa. Es decir, la edad de la madre y su raza no deberían ser influenciadas por la cantidad de trabajo que realiza una mujer. Al ser variables exógenas y no estar relacionadas con la cantidad de trabajo, es poco probable que su inclusión cambie significativamente los resultados de la regresión y la significancia del parámetro estimado para `morekids`.

Por otra parte, si esas variables estuvieran correlacionadas con el error, incluirlas en el modelo podría alterar el coeficiente de `morekids` y, por lo tanto, podrían cambiar los resultados de la regresión

Realizando la regresión lineal múltiple:

```
rg2 <- lm(weeksm1 ~ morekids + agem1 + black + hispan + othrace, data = fertility_1)
stargazer(rgIV,rg2,type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               weeksm1
##                               instrumental      OLS
##                               variable
##                               (1)              (2)
## -----
## morekids                -6.314***           -6.230***
##                          (1.275)            (0.088)
##
## agem1                    0.838***
##                          (0.013)
##
## black                    11.664***
##                          (0.192)
##
## hispan                   0.466***
##                          (0.179)
##
## othrace                  2.142***
##                          (0.203)
##
## Constant                21.421***           -4.835***
##                          (0.487)            (0.385)
## -----
## Observations              254,654           254,654
## R2                        0.014             0.044
## Adjusted R2               0.014             0.044
## Residual Std. Error 21.715 (df = 254652)    21.384 (df = 254648)
## F Statistic              2,330.779*** (df = 5; 254648)
## =====
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

En ambas regresiones, el parametro que acompaña la variable independiente morekids es alrededor de -6. Los resultados de la regresión IV y de la regresión lineal múltiple indica que la inclusión de las variables agem1, black, hispan y othrace no cambia significativamente el efecto de la fertilidad en la oferta laboral, representada por morekids, dado que se consideran exógenas, lo que significa que no están directamente influenciadas por la cantidad de trabajo. Por otra parte, al observar el parámetro  $\hat{\beta}_0$  en la regresión con las variables que consideramos exogenas, éste es -4.835, siendo imposible que, el hecho de que, cuando no tengas hijos y las demás variables sean cero, trabajes una cantidad de semanas negativas. Lo anterior puede generarse debido a la presencia de variables omitidas importantes o posible multicolinealidad entre las variables.