

Determinantes de la asignación de inversiones en planes de pensión: evidencia de mujeres maduras en EE.UU.

```
# Paquetes a utilizar:  
#| echo: false  
#| warning: false  
#| message: false  
library(tidyverse)
```

Warning: package 'tibble' was built under R version 4.5.1

Warning: package 'purrr' was built under R version 4.5.1

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
v dplyr      1.1.4      v readr      2.1.5  
v forcats    1.0.0      v stringr    1.5.1  
v ggplot2    3.5.2      v tibble     3.3.0  
v lubridate  1.9.4      v tidyr      1.3.1  
v purrr      1.1.0  
-- Conflicts ----- tidyverse_conflicts() --  
x dplyr::filter() masks stats::filter()  
x dplyr::lag()     masks stats::lag()  
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

Pregunta de investigación: ¿En qué medida los factores sociodemográficos y financieros explican la proporción invertida en acciones dentro de los planes de pensión?

Variable dependiente:

pctstck: Porcentaje invertido en acciones dentro del plan de pensión

Variables explicativas:

educ: Nivel educativo alcanzado

wealth89: Riqueza neta en 1989

age: Edad en años

married: Valor 1 si está casada

irain89: Valor 1 si tenía una cuenta IRA en 1989

Base de datos utilizada: pension

Introducción • Presenta el problema o fenómeno económico a estudiar. • Justifica su relevancia empírica o teórica. • Formula claramente la pregunta de investigación. • Expón de forma general tu hipótesis o lo que esperas encontrar. 2. Revisión de literatura • Resume brevemente estudios previos relevantes. • Enfócate en trabajos que hayan usado datos o modelos similares. • Identifica brechas que tu trabajo busca explorar. 1 • Usa citas correctamente en estilo APA. 3. Datos y metodología • Describe las fuentes de datos utilizadas y el período temporal. • Define cada variable (dependiente e independientes). • Presenta el modelo de regresión lineal múltiple propuesto (ecuación con notación formal). • Justifica la selección de variables. 4. Estimaciones y resultados • Presenta la tabla con resultados del modelo (coeficientes, errores estándar, R^2). • Interpreta los coeficientes principales en términos económicos. • Comenta sobre la significancia estadística y sentido económico de los resultados. • No incluyas aquí especulaciones o recomendaciones (eso va en la conclusión). 5. Conclusiones • Resume los hallazgos principales y su coherencia con la literatura revisada. • Expón posibles implicaciones para políticas públicas o futuras investigaciones. • Señala limitaciones del análisis (si aplica).

```
pension <- readxl::read_excel("pension.xlsx")
```

Descripción de las variables

La variable dependiente es una variable continua que representa el porcentaje de acciones en el plan de pensión. Esta variable es una variable limitada entre 0 y 100, ya que no se puede invertir más del 100% en acciones. Esta variable es truncada o limitada y no tiene una distribución normal, lo que puede afectar la validez de los resultados del modelo de regresión lineal.

Los modelos tobit son adecuados para este tipo de variables, ya que permiten modelar la relación entre la variable dependiente y las variables independientes, teniendo en cuenta el truncamiento de la variable dependiente.

Las variables explicativas son: - **educ:** Variable categórica que representa el nivel educativo alcanzado por la persona. Esta variable puede tener un efecto significativo en la proporción de

acciones en el plan de pensión, ya que las personas con mayor educación tienden a tener una mayor comprensión de los mercados financieros y, por lo tanto, pueden estar más dispuestas a invertir en acciones.

Datos y metodología

Los datos utilizados en este estudio provienen de la base de datos pension, que contiene información sobre mujeres maduras en Estados Unidos. La base de datos incluye variables sociodemográficas y financieras, así como la proporción de acciones en el plan de pensión.

La variables de intereses son las siguientes:

- **pctstck**: Porcentaje invertido en acciones dentro del plan de pensión (variable dependiente).
- **educ**: Nivel educativo alcanzado (variable independiente).
- **wealth89**: Riqueza neta en 1989 (variable independiente).
- **age**: Edad en años (variable independiente).
- **married**: Valor 1 si está casada (variable independiente).
- **irain89**: Valor 1 si tenía una cuenta IRA en 1989 (variable independiente).

Deb

Estimación del modelo de regresión lineal:

```
#Variable dependiente:  
summary(pension$pctstck)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	0.00	50.00	48.45	100.00	100.00

```
modelo <- lm(pctstck ~ educ + wealth89 + age + married + irain89, data = pension)  
stargazer::stargazer(modelo, type = "text", title = "Modelo de regresión lineal para pctstck")
```

Modelo de regresión lineal para pctstck

```
=====
                        Dependent variable:
                        -----
                        pctstck
-----
educ                    0.542
                        (1.122)

wealth89                -0.004
                        (0.012)

age                    -1.458**
                        (0.701)

married                 5.004
                        (6.920)

irain89                -6.307
                        (5.961)

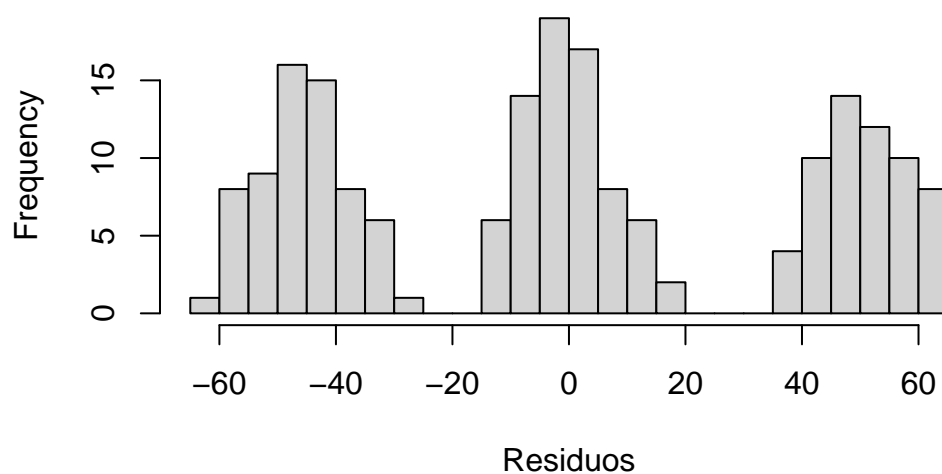
Constant                129.615***
                        (45.057)

-----
Observations            194
R2                      0.037
Adjusted R2             0.011
Residual Std. Error     39.503 (df = 188)
F Statistic             1.432 (df = 5; 188)
=====
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

Comprobando la normalidad de los residuos:

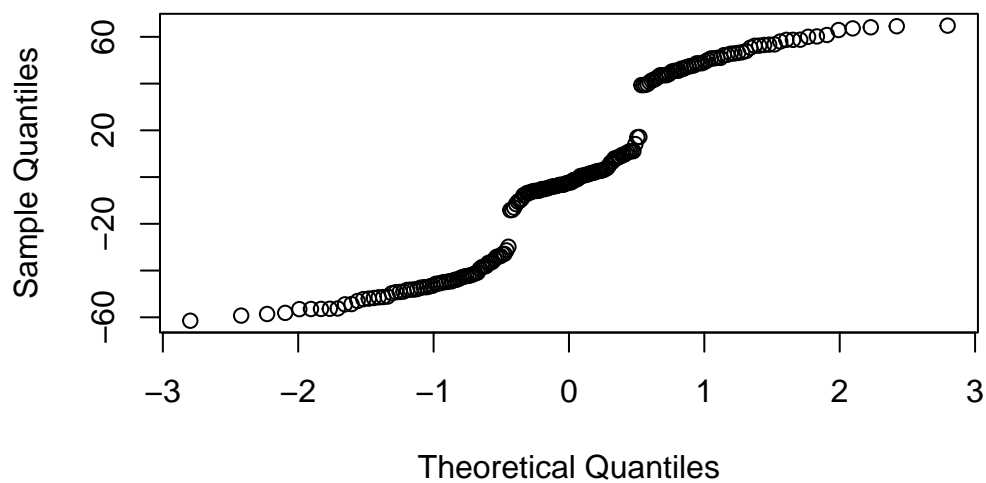
```
# Histograma de los residuos
hist(residuals(modelo), main = "Histograma de los residuos", xlab = "Residuos", breaks = 20)
```

Histograma de los residuos



```
# QQ-plot de los residuos  
qqnorm(residuals(modelo))
```

Normal Q-Q Plot



Análisis de heterocedasticidad

Planteamiento de la hipótesis nula: La varianza de los errores es constante (homocedasticidad).

$$H_0 : Var(u_i) = \sigma^2 \quad (\text{homocedasticidad})$$

$$H_a : Var(u_i) \neq \sigma^2 \quad (\text{heterocedasticidad})$$

```
# Prueba de Breusch-Pagan  
library(lmtest)
```

Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

as.Date, as.Date.numeric

```
library(sandwich)  
bptest(modelo)
```

studentized Breusch-Pagan test

data: modelo
BP = 7.7405, df = 5, p-value = 0.1711

```
# Prueba de White  
white_test <- bptest(modelo, ~ fitted(modelo) + I(fitted(modelo)^2))  
white_test
```

studentized Breusch-Pagan test

data: modelo
BP = 4.1471, df = 2, p-value = 0.1257

Prueba de Breusch-Pagan

El p-value es de 0.1711 lo que indica que no se rechaza la hipótesis nula de homocedasticidad. Concluyendo que al nivel de significancia del 5% no hay evidencia de heterocedasticidad en el modelo.

Prueba de White

En este test el p-value es de 0.1257 lo que también indica que no se rechaza la hipótesis nula de homocedasticidad. Por lo tanto, al igual que en la prueba de Breusch-Pagan, no hay evidencia de heterocedasticidad en el modelo.

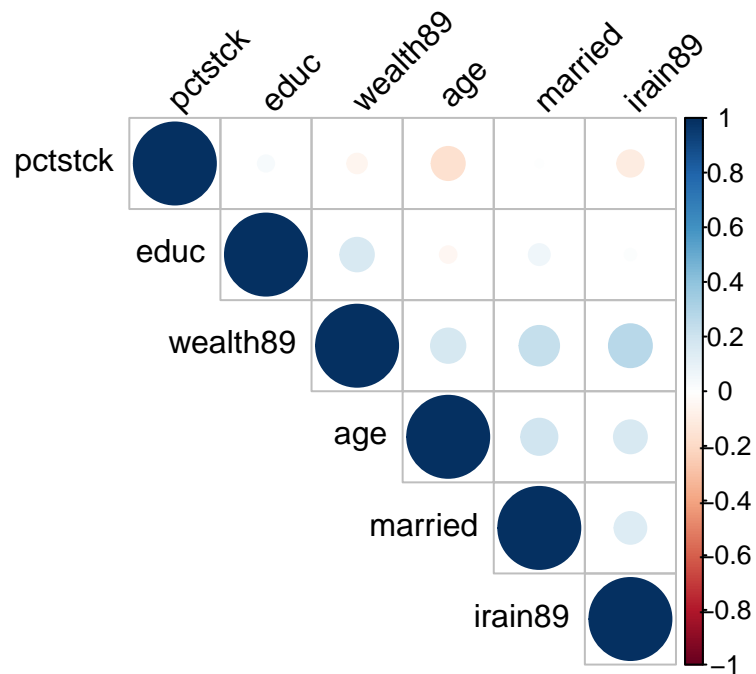
Análisis de multicolinealidad

Supuesto:

```
# Matriz de correlación
cor_matrix <- cor(pension[, c("pctstck", "educ", "wealth89", "age", "married", "irain89")], t
# Visualización de la matriz de correlación
library(corrplot)
```

corrplot 0.95 loaded

```
corrplot(cor_matrix, method = "circle", type = "upper", tl.col = "black", tl.srt = 45)
```



```
# VIF para detectar multicolinealidad
library(car)
```

Loading required package: carData

Attaching package: 'car'

The following object is masked from 'package:dplyr':

recode

The following object is masked from 'package:purrr':

some

```
vif_values <- vif(modelo)
vif_values
```

```
educ wealth89 age married irain89
1.036298 1.172165 1.077044 1.092894 1.103397
```


Anexos:

Gráficos de las variables y validez de los resultados

```
# plot(pension$ed)
```