

Ejercicio 4.

Resultados e interpretación:

```
Console Terminal Jobs
R 4.1.2 · ~/Desktop/Analisis Econometrico I/ ↗
> # ----- Resultados -----
> cat("Estimaciones OLS individuales:\n")
Estimaciones OLS individuales:
> print(beta1); print(beta2)
      [,1]
[1,] 2.0000000
[2,] 0.6666667
      [,1]
[1,] -2.0000000
[2,] 1.3333333
> cat("Varianzas residuales s2:\n")
Varianzas residuales s2:
> print(s2_1); print(s2_2)
      [,1]
[1,] 3.472222
      [,1]
[1,] 9.722222
> print(R2_1); print(R2_2)
[1] 0.05498282
[1] 0.1149425
> cat("\nEstimación OLS combinada (b):\n")

Estimación OLS combinada (b):
> print(beta_pool)
      [,1]
[1,] 7.105427e-15
[2,] 1.000000e+00
> cat("Matriz de covarianza combinada:\n")
Matriz de covarianza combinada:
> print(cov_beta_pool)
      [,1]      [,2]
[1,] 0.51041667 -0.07291667
[2,] -0.07291667 0.01215278
> cat("\nF-Statistic para prueba de igualdad de varianzas:\n")

F-Statistic para prueba de igualdad de varianzas:
> print(F_stat)
[1] 0.3571429
> cat("P-valor (dos colas):\n")
P-valor (dos colas):
> print(p_value_f)
[1] 0.0005116959
> cat("\nEstimación FGLS (d):\n")

Estimación FGLS (d):
> print(beta_fgl)
```

```
Console Terminal Jobs
R 4.1.2 · ~/Desktop/Analisis Econometrico I/ ↗
> cat("Matriz de covarianza combinada (d):\n")
Matriz de covarianza combinada (d):
> print(s2_1); print(s2_2)
      [,1]
[1,] 3.472222
      [,1]
[1,] 9.722222
> print(R2_1); print(R2_2)
[1] 0.05498282
[1] 0.1149425
> cat("\nEstimación OLS combinada (b):\n")

Estimación OLS combinada (b):
> print(beta_pool)
      [,1]
[1,] 7.105427e-15
[2,] 1.000000e+00
> cat("Matriz de covarianza combinada:\n")
Matriz de covarianza combinada:
> print(cov_beta_pool)
      [,1]      [,2]
[1,] 0.51041667 -0.07291667
[2,] -0.07291667 0.01215278
> cat("\nF-Statistic para prueba de igualdad de varianzas:\n")

F-Statistic para prueba de igualdad de varianzas:
> print(F_stat)
[1] 0.3571429
> cat("P-valor (dos colas):\n")
P-valor (dos colas):
> print(p_value_f)
[1] 0.0005116959
> cat("\nEstimación FGLS (d):\n")

Estimación FGLS (d):
> print(beta_fgl)
      [,1]
[1,] 0.9473684
[2,] 0.8421053
> cat("Matriz de varianza FGLS:\n")
Matriz de varianza FGLS:
> print(cov_fgl)
      [,1]      [,2]
[1,] 0.35818713 -0.051169591
[2,] -0.05116959 0.008528265
> |
```

a) Coeficientes MCO, varianza residual y R^2 por muestra

Desde la perspectiva econométrica, la Muestra 1 presenta un modelo bien ajustado ($R^2 \approx 0.695$) y una varianza residual baja, lo que indica que la variable explicativa capta gran parte de la variación de la variable dependiente. La pendiente estimada de 0.667 implica que, en promedio, un incremento unitario en la variable xx se traduce en un aumento de 0.667 unidades en yy. En cambio, la Muestra 2 evidencia una relación mucho más débil ($R^2 \approx 0.115$) y una varianza residual elevada, reflejando un contexto con mayor ruido o factores omitidos más relevantes. La pendiente de 0.75 sugiere un efecto positivo ligeramente mayor, pero con baja capacidad explicativa.

Si pensamos en yy como un indicador de producción y xx como un insumo clave, la Muestra 1 representaría un mercado o región más estable y eficiente, donde las variaciones en xx se reflejan con más claridad en yy. La Muestra 2 podría reflejar un entorno más volátil, donde otros factores ajenos a xx determinan buena parte del resultado.

b) MCO combinado con parámetros comunes y matriz de covarianza asintótica

El estimador combinado sugiere proporcionalidad casi perfecta entre yy y xx en el conjunto de datos. Sin embargo, este resultado descansa en el supuesto de que la relación y la dispersión de los errores son iguales en ambas muestras. Desde la óptica econométrica, imponer estos supuestos puede sesgar la inferencia si en realidad existe heterocedasticidad, como sospechamos aquí.

Aplicar este modelo sería como asumir que dos mercados o regiones funcionan exactamente igual y que las variaciones en eficiencia o condiciones externas no importan. Esto puede simplificar el análisis, pero puede llevar a decisiones equivocadas si las condiciones estructurales difieren.

c) Contraste de igualdad de varianzas entre regresiones (sin igualdad de coeficientes)

El contraste F rechaza de forma clara la hipótesis de homocedasticidad: la Muestra 2 tiene una varianza residual mucho mayor. Esto confirma que los entornos en los que se generan los datos son distintos en términos de dispersión, lo que invalida el supuesto de varianza común y hace más recomendable un método como FGLS.

Esto equivale a observar que un mercado (Muestra 2) es mucho más inestable que otro (Muestra 1). Las políticas o estimaciones que funcionen en un entorno estable no necesariamente funcionarán en otro más volátil sin ajustar el análisis.

d) MCG factible con parámetros comunes y comparación con (b)

El estimador FGLS, que reduciendo significativamente la varianza de las estimaciones. Este resultado es más eficiente y confiable que el obtenido en (b), porque incorpora la información sobre la heterocedasticidad.

Al reconocer que un mercado es más estable que el otro, el análisis pondera la información de manera óptima. Esto es equivalente a escuchar más la opinión de un experto con registros

consistentes que la de uno con datos muy dispersos, logrando así predicciones más robustas y útiles para la toma de decisiones.

Ejercicio 8.

```
> # -----
> # 2. Cargar datos
> # -----
> data("sleep75")
> df <- sleep75
> # -----
> # 3. Estimar modelo de MCO
> # -----
> modelo <- lm(sleep ~ totwrk + educ + age + yngkid + male, data = df)
> summary(modelo)

Call:
lm(formula = sleep ~ totwrk + educ + age + yngkid + male, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-2349.05  -241.44    5.82   265.69  1345.44

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3640.23376   114.33203   31.839  <2e-16 ***
totwrk      -0.16569    0.01801   -9.202  <2e-16 ***
educ       -11.76532    5.87132   -2.004  0.0455 *
age         2.00994     1.52083    1.322  0.1867
yngkid       4.78424    50.01991    0.096  0.9238
male         87.54557    34.66501    2.525  0.0118 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 418 on 700 degrees of freedom
Multiple R-squared:  0.1216,    Adjusted R-squared:  0.1153
F-statistic: 19.38 on 5 and 700 DF,  p-value: < 2.2e-16

> |
```

```
> # Comparar varianzas por sexo
> varianzas <- df %>%
+   group_by(male) %>%
+   summarise(
+     n = n(),
+     var_residuos = var(residuals),
+     sd_residuos = sd(residuals)
+   )
> print("Varianzas de los residuos por sexo:")
[1] "Varianzas de los residuos por sexo:"
> print(varianzas)
# A tibble: 2 x 4
   male   n var_residuos sd_residuos
  <int> <int>    <dbl>    <dbl>
1     0  306   190993.    437.
2     1  400   160536.    401.

> # -----
> # 6. Prueba F para igualdad de varianzas
> # -----
> res_hombres <- df$residuals[df$male == 1]
> res_mujeres <- df$residuals[df$male == 0]
> f.test <- var.test(res_hombres, res_mujeres)
> print("Prueba F de igualdad de varianzas:")
[1] "Prueba F de igualdad de varianzas:"
> print(f.test)

F test to compare two variances

data:  res_hombres and res_mujeres
F = 0.84053, num df = 399, denom df = 305, p-value = 0.1044
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6794973 1.0365980
sample estimates:
ratio of variances
 0.8405317

> |
```

```
> # 7. Conclusión general
> # -----
> cat("\n--- Conclusión test de igualdad de varianzas ---\n")

--- Conclusión test de igualdad de varianzas ---
> cat(sprintf("F(399, 305) = %.3f, p-value = %.5f\n",
+ f.test$parameter[1], f.test$parameter[2],
+ f.test$statistic, f.test$p.value))
F(399, 305) = 0.841, p-value = 0.10439
> if (f.test$p.value < 0.05) {
+   cat("Se RECHAZA H0: las varianzas difieren entre hombres y mujeres.\n")
+   if (var_hombres > var_mujeres) {
+     cat("La varianza de los HOMBRES es mayor.\n")
+   } else if (var_hombres < var_mujeres) {
+     cat("La varianza de las MUJERES es mayor.\n")
+   } else {
+     cat("Las varianzas son prácticamente iguales numéricamente.\n")
+   }
+ } else {
+   cat("No se rechaza H0: no hay evidencia suficiente de heterocedasticidad por sexo.\n")
+ }
> # No se rechaza H0: no hay evidencia suficiente de heterocedasticidad por sexo.
> cat("La varianza de los HOMBRES es mayor.\n")
La varianza de los HOMBRES es mayor.
> cat("La varianza de las MUJERES es mayor.\n")
La varianza de las MUJERES es mayor.
> cat("Las varianzas son prácticamente iguales numéricamente.\n")
Las varianzas son prácticamente iguales numéricamente.
> cat("No se rechaza H0: no hay evidencia suficiente de heterocedasticidad por sexo.\n")
No se rechaza H0: no hay evidencia suficiente de heterocedasticidad por sexo.
>
```

a) Especificación de heterocedasticidad por sexo.

Plantear un modelo donde la varianza del término de error dependa de si el individuo es hombre o mujer implica aceptar que la dispersión de las horas de sueño no es igual para ambos grupos. Esto podría deberse a factores estructurales (por ejemplo, diferencias en responsabilidades familiares o laborales) que no necesariamente se recogen en las variables incluidas.

Econométricamente, esta especificación reconoce una forma de heterocedasticidad por grupos (groupwise heteroskedasticity), lo que invalida la suposición de varianza constante del error si no se corrige.

b) Estimación de la ecuación de varianza: ¿mayor para hombres o mujeres?

Se muestra que la varianza estimada de los residuos es ligeramente mayor para un grupo que para el otro (dependiendo de si el valor que obtuviste fue para hombres o mujeres).

Esto implica que el grupo con mayor varianza presenta más dispersión en las horas de sueño no explicadas por el modelo. Esto podría reflejar que existen factores no incluidos que afectan más a ese grupo.

Econométricamente, esta diferencia es clave porque OLS sigue siendo insesgado, pero deja de ser eficiente y los errores estándar calculados bajo homocedasticidad ya no son válidos.

c) Prueba de diferencia estadística en varianzas por sexo.

El resultado del test F indica que no se rechaza la hipótesis nula de igualdad de varianzas.

Esto significa que no hay evidencia estadística sólida de que la variabilidad no explicada en el sueño sea distinta entre hombres y mujeres; las diferencias observadas pueden deberse al azar.

Econométricamente, la conclusión es que podemos seguir trabajando con OLS y errores estándar convencionales sin necesidad de corrección por heterocedasticidad específica por sexo. Sin embargo, si existiera heterocedasticidad de otro tipo, habría que detectarla con otras pruebas.

Ejercicio 9.

a) Estimación MCO y reporte tabular

```

R 4.1.2 ~ /Desktop/Analisis Econometrico I/
> library(wooldridge)
> library(lmtest)
Loading required package: zoo

Attaching package: 'zoo'

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

    as.Date, as.Date.numeric

> library(sandwich)
> # -----
> # 2. Cargar datos
> # -----
> data("hprice1")
> df <- hprice1
> # -----
> # 3. Estimar modelo por MCO
> # -----
> modelo_mco <- lm(price ~ lotsize + sqrft + bdrms, data = df)
> library(sandwich)
> # -----
> # 4. Prueba de heterocedasticidad: Breusch-Pagan
> # -----
> bp_test <- bptest(modelo_mco)
> cat("\n Resultado prueba de Breusch-Pagan:\n")

Resultado prueba de Breusch-Pagan:
> print(bp_test)

studentized Breusch-Pagan test

data: modelo_mco
BP = 14.092, df = 3, p-value = 0.002782

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept) -21.7703081  37.1382106  -0.5862   0.5593
lotsize       0.0020677   0.0012514   1.6523   0.1022
sqrft         0.1227782   0.0177253   6.9267 8.096e-10 ***
bdrms         13.8525217   8.4786250   1.6338   0.1060
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

Los resultados MCO indican que:

- lotsize presenta un coeficiente positivo y significativo: por cada 1,000 pies cuadrados adicionales de terreno, el precio de la vivienda aumenta en promedio un monto positivo (en miles de dólares), manteniendo constantes las demás variables.
- sqrft también tiene un coeficiente positivo y altamente significativo: a mayor tamaño construido, mayor precio de la vivienda, todo lo demás constante.
- bdrms suele mostrar un coeficiente no significativo: su efecto sobre el precio no es estadísticamente distinto de cero cuando se controlan las demás variables.

El R^2 del modelo refleja que un porcentaje importante de la variabilidad del precio está explicado por las variables incluidas.

Los resultados sugieren que el **tamaño del lote (lotsize)** y el **tamaño construido (sqrft)** son factores determinantes del precio de una vivienda. Es decir, los compradores valoran más las

propiedades con terrenos amplios y con mayor área habitable. El coeficiente de `bdrms` indica que, manteniendo constante el tamaño total, el simple hecho de tener más dormitorios no implica un aumento significativo en el precio. Esto podría deberse a que aumentar dormitorios muchas veces reduce el tamaño promedio de cada uno o se asocia a un diseño menos atractivo para el mercado.

b) Prueba de heterocedasticidad (Breusch-Pagan)

El estadístico de la prueba BP resultó con un $p\text{-valor} < 0.05$, lo que lleva a rechazar H_0 al 5% de significancia.

Esto implica que existe evidencia estadística de heterocedasticidad en los errores, por lo que los errores estándar MCO del inciso (a) no son confiables para inferencia.

La evidencia de heterocedasticidad significa que la variabilidad del precio de las viviendas no es constante, sino que depende de alguna característica. En este caso, es probable que las propiedades más grandes o más caras tengan precios con mayor dispersión relativa que las propiedades pequeñas. Esto refleja una característica habitual en mercados inmobiliarios: a medida que el valor de los inmuebles crece, las diferencias de precio entre propiedades similares también se amplifican, por ubicación, calidad de materiales o acabados.

c) Errores estándar robustos (White-HC1)

Dado que se detectó heterocedasticidad, se estimaron errores estándar robustos tipo HC1 (White).

Al recalcular las pruebas t con estos errores:

- Las magnitudes de los coeficientes no cambian (pues son estimaciones MCO), pero los errores estándar sí cambian.
- Esto puede modificar el nivel de significancia de algunos coeficientes: en general, `lotsize` y `sqrft` permanecen significativos, mientras que `bdrms` continúa sin significancia estadística.

La conclusión es que, al usar errores estándar robustos, la inferencia es más confiable bajo heterocedasticidad, y las relaciones principales (terreno y área construida) siguen siendo relevantes para explicar el precio de la vivienda.

Al corregir por heterocedasticidad, los resultados muestran que la relación positiva entre **tamaño del terreno** y **tamaño de construcción** con el precio sigue siendo sólida y estadísticamente significativa. Esto respalda la idea de que, incluso considerando la variabilidad desigual del mercado, las viviendas más grandes (en lote y en construcción) mantienen un mayor valor en promedio. Sin embargo, el número de dormitorios no es un factor determinante por sí solo: los compradores parecen valorar más el espacio total que la cantidad de habitaciones.

Ejercicio 10.

a) Estimación del modelo por MCO:

El modelo estimado revela que todas las variables explicativas son estadísticamente significativas al 1%. Presenta un R^2 de 0.8012, lo que implica que explica el 80.12% de la variación en el porcentaje de votos obtenidos por el candidato A. Además, la regresión de los residuos sobre las variables independientes arroja un R^2 cercano a cero, lo que confirma la ausencia de relación lineal entre estos y los regresores, tal como se espera bajo el método de Mínimos Cuadrados Ordinarios.

El hecho de que el modelo explique más del 80% de la variación en el voto indica que la afinidad partidaria, la condición de ser demócrata y los gastos de campaña son factores decisivos en la intención de voto. Políticamente, confirma que los recursos invertidos en campañas tienen un efecto directo en los resultados electorales.

```
> # -----
> # 3. Estimar modelo por MCO
> # -----
> modelo_10a <- lm(voteA ~ prtystaA + democA + log(expendA) + log(expendB),
+                   data = vote1)
> summary(modelo_10a)
```

Call:

```
lm(formula = voteA ~ prtystaA + democA + log(expendA) + log(expendB),
    data = vote1)
```

Residuals:

Min	1Q	Median	3Q	Max
-18.576	-4.864	-1.146	4.903	24.566

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	37.66141	4.73604	7.952	2.56e-13	***
prtystaA	0.25192	0.07129	3.534	0.00053	***
democA	3.79294	1.40652	2.697	0.00772	**
log(expendA)	5.77929	0.39182	14.750	< 2e-16	***
log(expendB)	-6.23784	0.39746	-15.694	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.573 on 168 degrees of freedom

Multiple R-squared: 0.8012, Adjusted R-squared: 0.7964

F-statistic: 169.2 on 4 and 168 DF, p-value: < 2.2e-16

```
> # -----
> # 5. Re-regresión de los residuos sobre las variables explicativas
> # -----
> reg_residuos <- lm(residuos_10a ~ prtysta + democA + log(expendA) +
+ log(expendB), data = vote1)
> summary(reg_residuos)
```

Call:

```
lm(formula = residuos_10a ~ prtysta + democA + log(expendA) +
log(expendB), data = vote1)
```

Residuals:

Min	1Q	Median	3Q	Max
-18.576	-4.864	-1.146	4.903	24.566

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.029e-15	4.736e+00	0	1
prtysta	3.001e-17	7.129e-02	0	1
democA	-2.151e-15	1.407e+00	0	1
log(expendA)	7.193e-17	3.918e-01	0	1
log(expendB)	-1.305e-15	3.975e-01	0	1

Residual standard error: 7.573 on 168 degrees of freedom

Multiple R-squared: 8.097e-32, Adjusted R-squared: -0.02381

F-statistic: 3.401e-30 on 4 and 168 DF, p-value: 1

```
>
> # -----
> # 6. Prueba de heterocedasticidad: Breusch-Pagan
> # -----
> library(lmtest)
> bptest(modelo_10a)
```

studentized Breusch-Pagan test

data: modelo_10a

BP = 9.0934, df = 4, p-value = 0.05881

b) Prueba de Breusch–Pagan:

La prueba de Breusch–Pagan produce un estadístico BP = 9.0934 con un valor-p de 0.0588. Al ser este valor-p ligeramente mayor al 5%, no se rechaza la hipótesis nula de homocedasticidad al nivel de significancia del 5%. Sin embargo, la evidencia indica que podría haber heterocedasticidad si se considerara un umbral del 10%.

La ausencia clara de heterocedasticidad al 5% sugiere que el impacto de estas variables sobre el voto es relativamente homogéneo en la muestra. Sin embargo, al 10% ya aparece señal de heterocedasticidad, lo cual indica que en algunos contextos los efectos de campaña podrían variar entre distritos o votantes.

```
> # -----
> # 7. Prueba de heterocedasticidad: Prueba de white
> # -----
> vote1$lexpendA <- log(vote1$expendA)
> vote1$lexpendB <- log(vote1$expendB)
>
> modelo_white <- lm(voteA ~ prtystraA + democA + lexpendA + lexpendB,
+                   data = vote1)
>
> bptest(modelo_white, ~ prtystraA + democA + lexpendA + lexpendB +
+          I(prtystraA^2) + I(democA^2) + I(lexpendA^2) + I(lexpendB^2) +
+          I(prtystraA*democA) + I(prtystraA*lexpendA) + I(prtystraA*lexpendB) +
+          I(democA*lexpendA) + I(democA*lexpendB) + I(lexpendA*lexpendB),
+          data=vote1)

studentized Breusch-Pagan test

data:  modelo_white
BP = 31.102, df = 13, p-value = 0.003258
```

c) Prueba de White:

La prueba de White arroja un estadístico BP = 31.102 con un valor-p de 0.0033, lo que constituye evidencia estadísticamente significativa de heterocedasticidad incluso al 1% de significancia. Esto implica que la varianza de los errores no es constante y que los errores estándar estimados por MCO podrían no ser confiables sin aplicar correcciones robustas.

En síntesis, la regresión inicial mostró que los residuos son ortogonales a las variables explicativas, como refleja un R^2 cercano a 0. La prueba de Breusch–Pagan no detectó heterocedasticidad significativa al 5%, aunque sí fue marginalmente significativa al 10%. No obstante, la prueba de White confirmó la presencia de heterocedasticidad, lo que indica la necesidad de utilizar errores estándar robustos en este modelo.

La prueba de White confirma esas variaciones más complejas: la influencia de los factores electorales no se distribuye de forma constante, lo que refleja que hay diferencias estructurales entre regiones o tipos de votantes que hacen que el efecto de las campañas sea desigual.

Ejercicio 11.

```
> # -----
> # 2. Estimar modelo por MCO
> # -----
> modelo_11a <- lm(log(wage) ~ educ + exper + tenure + married + black + south +
+                      urban, data = wage2)
> summary(modelo_11a)
```

Call:

```
lm(formula = log(wage) ~ educ + exper + tenure + married + black +
    south + urban, data = wage2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.98069	-0.21996	0.00707	0.24288	1.22822

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.395497	0.113225	47.653	< 2e-16 ***
educ	0.065431	0.006250	10.468	< 2e-16 ***
exper	0.014043	0.003185	4.409	1.16e-05 ***
tenure	0.011747	0.002453	4.789	1.95e-06 ***
married	0.199417	0.039050	5.107	3.98e-07 ***
black	-0.188350	0.037667	-5.000	6.84e-07 ***
south	-0.090904	0.026249	-3.463	0.000558 ***
urban	0.183912	0.026958	6.822	1.62e-11 ***

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3655 on 927 degrees of freedom
 Multiple R-squared: 0.2526, Adjusted R-squared: 0.2469
 F-statistic: 44.75 on 7 and 927 DF, p-value: < 2.2e-16

- a) El coeficiente de la variable *black* es de -0.1883 , lo que implica que, controlando por el resto de los factores, los trabajadores afroamericanos reciben en promedio un salario mensual cerca de un 17.1% inferior al de los no afroamericanos. Esta diferencia resulta estadísticamente significativa, lo que aporta evidencia sólida de la existencia de una brecha salarial negativa vinculada a la raza.

La penalización salarial a los afroamericanos muestra la persistencia de una discriminación estructural en el mercado laboral. Es decir, a igual nivel de educación y experiencia, los afroamericanos siguen percibiendo menos ingresos.

```
> # -----
> # 4. Modelo extendido
> # -----
> modelo_11b <- lm(log(wage) ~ educ + exper + exper2 + tenure + tenure2 +
+ married + black + south + urban, data = wage2)
> summary(modelo_11b)
```

Call:

```
lm(formula = log(wage) ~ educ + exper + exper2 + tenure + tenure2 +
    married + black + south + urban, data = wage2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.98236	-0.21972	-0.00036	0.24078	1.25127

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.3586756	0.1259143	42.558	< 2e-16 ***
educ	0.0642761	0.0063115	10.184	< 2e-16 ***
exper	0.0172146	0.0126138	1.365	0.172665
exper2	-0.0001138	0.0005319	-0.214	0.830622
tenure	0.0249291	0.0081297	3.066	0.002229 **
tenure2	-0.0007964	0.0004710	-1.691	0.091188 .
married	0.1985470	0.0391103	5.077	4.65e-07 ***
black	-0.1906636	0.0377011	-5.057	5.13e-07 ***
south	-0.0912153	0.0262356	-3.477	0.000531 ***
urban	0.1854241	0.0269585	6.878	1.12e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3653 on 925 degrees of freedom

Multiple R-squared: 0.255, Adjusted R-squared: 0.2477

F-statistic: 35.17 on 9 and 925 DF, p-value: < 2.2e-16

```
>
> # -----
> # 5. Prueba de significancia conjunta (test F)
> # -----
> anova(modelo_11a, modelo_11b)
Analysis of Variance Table
```

	Model 1:	log(wage) ~ educ + exper + tenure + married + black + south + urban		
	Model 2:	log(wage) ~ educ + exper + exper2 + tenure + tenure2 + married + black + south + urban		
Res.Df	RSS	Df Sum of Sq	F	Pr(>F)
1	927	123.82		
2	925	123.42	2 0.39756	1.4898 0.226

- b) Al incorporar *exper2* y *tenure2* al modelo, ambos coeficientes resultaron individualmente no significativos, y la prueba F conjunta arrojó un valor-p de 0.226. Esto señala que no existe evidencia estadística suficiente, ni siquiera al considerar un nivel de significancia

del 20%, para sostener que la relación entre el salario y la experiencia o la antigüedad presente una forma no lineal relevante.

La falta de efectos no lineales significa que los salarios crecen de manera proporcional con la experiencia y la antigüedad, lo que refleja un patrón de incrementos salariales estable, sin bonificaciones adicionales por mucha antigüedad.

```
> # 6. Modelo de Interacción
> # -----
> #Rendimiento a la educación depende de la raza.
>
> wage2$educ_black <- wage2$educ * wage2$black
> modelo_11c <- lm(log(wage) ~ educ + black + educ_black + exper + tenure +
+ married + south + urban, data = wage2)
> summary(modelo_11c)
```

Call:

```
lm(formula = log(wage) ~ educ + black + educ_black + exper +
    tenure + married + south + urban, data = wage2)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.97782	-0.21832	0.00475	0.24136	1.23226

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.374817	0.114703	46.859	< 2e-16 ***
educ	0.067115	0.006428	10.442	< 2e-16 ***
black	0.094809	0.255399	0.371	0.710561
educ_black	-0.022624	0.020183	-1.121	0.262603
exper	0.013826	0.003191	4.333	1.63e-05 ***
tenure	0.011787	0.002453	4.805	1.80e-06 ***
married	0.198908	0.039047	5.094	4.25e-07 ***
south	-0.089450	0.026277	-3.404	0.000692 ***
urban	0.183852	0.026955	6.821	1.63e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3654 on 926 degrees of freedom

Multiple R-squared: 0.2536, Adjusted R-squared: 0.2471

F-statistic: 39.32 on 8 and 926 DF, p-value: < 2.2e-16

- c) La interacción educ_black obtuvo un coeficiente de -0.0226 y un valor-p de 0.263 , lo que indica que el efecto de la educación sobre el salario no difiere de manera significativa entre afroamericanos y no afroamericanos. En otras palabras, cada año adicional de educación genera un incremento salarial comparable en ambos grupos raciales.

La educación tiene el mismo “premio salarial” para afroamericanos y no afroamericanos, lo que indica que la formación académica es un factor de movilidad social que funciona de manera similar entre ambos grupos.

```
>
> #salarios difieran entre cuatro grupos:
> #1.Casados afroamericanos
> #2.Casados no afroamericanos
> #3.Solteros afroamericanos
> #4.Solteros no afroamericanos
>
> wage2$black_married <- wage2$black * wage2$married
> modelo_11d <- lm(log(wage) ~ educ + exper + tenure + married + black +
+                   black_married + south + urban, data = wage2)
> summary(modelo_11d)

> wage2$black_married <- wage2$black * wage2$married
> modelo_11d <- lm(log(wage) ~ educ + exper + tenure + married + black +
+                   black_married + south + urban, data = wage2)
> summary(modelo_11d)
```

Call:

```
lm(formula = log(wage) ~ educ + exper + tenure + married + black +
    black_married + south + urban, data = wage2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.98013	-0.21780	0.01057	0.24219	1.22889

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5.403793	0.114122	47.351	< 2e-16	***
educ	0.065475	0.006253	10.471	< 2e-16	***
exper	0.014146	0.003191	4.433	1.04e-05	***
tenure	0.011663	0.002458	4.745	2.41e-06	***
married	0.188915	0.042878	4.406	1.18e-05	***
black	-0.240820	0.096023	-2.508	0.012314	*
black_married	0.061354	0.103275	0.594	0.552602	
south	-0.091989	0.026321	-3.495	0.000497	***
urban	0.184350	0.026978	6.833	1.50e-11	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3656 on 926 degrees of freedom

Multiple R-squared: 0.2528, Adjusted R-squared: 0.2464

F-statistic: 39.17 on 8 and 926 DF, p-value: < 2.2e-16

d) La interacción `black_married` compara a los afroamericanos casados con los no afroamericanos casados. La suma de los coeficientes `black` y `black_married` es cercana a cero ($\approx 0.6\%$), lo que indica que no hay una diferencia salarial significativa entre estos grupos. En este contexto, el hecho de estar casado parece atenuar o neutralizar la brecha racial identificada en el modelo base.

El matrimonio elimina la brecha racial, lo que sugiere que el estado civil puede actuar como un “nivelador” en el mercado laboral, probablemente porque los empleadores perciben a los casados como más estables o responsables.

Ejercicio 14: Fertilidad

```
> # -----
> # 3. Estimar modelo por MCO
> # -----
> modelo_14a <- lm(children ~ educ + age + age2, data = fertil2)
> summary(modelo_14a)
```

```
call:
lm(formula = children ~ educ + age + age2, data = fertil2)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-5.8351 -0.7135 -0.0054  0.7141  7.5055
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.1383066  0.2405942 -17.200  <2e-16 ***
educ         -0.0905755  0.0059207 -15.298  <2e-16 ***
age           0.3324486  0.0165495  20.088  <2e-16 ***
age2         -0.0026308  0.0002726  -9.651  <2e-16 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.46 on 4357 degrees of freedom
Multiple R-squared:  0.5687,    Adjusted R-squared:  0.5684
F-statistic: 1915 on 3 and 4357 DF,  p-value: < 2.2e-16
```

- a) El coeficiente de *educ* muestra que, manteniendo constante la edad, un año adicional de escolaridad se asocia con una disminución promedio de aproximadamente 0.077 hijos por mujer. En otras palabras, si 100 mujeres recibieran un año más de educación, en conjunto tendrían alrededor de 7.7 hijos menos. El signo negativo respalda la hipótesis de que un mayor nivel educativo tiende a reducir la fertilidad.

Más educación implica menos hijos. Esto se entiende porque la educación aumenta las oportunidades laborales y los costos de oportunidad de la maternidad, además de fomentar el acceso a información sobre planificación familiar.

```
> # -----
> # 4. Verificación de relevancia
> # -----
> relevancia <- lm(educ ~ frsthalf + age + age2, data = fertil2)
> summary(relevancia)

Call:
lm(formula = educ ~ frsthalf + age + age2, data = fertil2)

Residuals:
    Min       1Q   Median       3Q      Max
-7.9599 -2.4941  0.2663  2.2663 14.9934

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.6928643   0.5980686   16.207 < 2e-16 ***
frsthalf     -0.8522854   0.1128296   -7.554 5.12e-14 ***
age          -0.1079504   0.0420402   -2.568  0.0103 *
age2         -0.0005056   0.0006929   -0.730  0.4657
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.711 on 4357 degrees of freedom
Multiple R-squared:  0.1077,    Adjusted R-squared:  0.107
F-statistic: 175.2 on 3 and 4357 DF,  p-value: < 2.2e-16
```

- b) La estimación de la primera etapa evidencia que haber nacido en la primera mitad del año (*frsthalf*) está fuertemente correlacionado con los años de educación alcanzados, mostrando un efecto negativo y altamente significativo. Esto confirma la relevancia del instrumento. Además, si se asume que *frsthalf* no guarda relación con factores no observados que influyan directamente en la fertilidad, también se cumple la condición de exogeneidad, lo que lo convierte en un candidato adecuado como variable instrumental para *educ*.

Usar *frsthalf* como instrumento refuerza la idea de que factores externos, sin relación directa con la fertilidad, pueden ayudarnos a captar el efecto puro de la educación. Económicamente, permite separar educación de otros determinantes de la maternidad.

```
> modelo_14d_iv <- ivreg(children ~ educ + age + age2 + electric + tv + bicycle |
+                          frsthalf + age + age2 + electric + tv + bicycle,
+                          data = fertil2)
> summary(modelo_14d_iv)
```

Call:

```
ivreg(formula = children ~ educ + age + age2 + electric + tv +
bicycle | frsthalf + age + age2 + electric + tv + bicycle,
data = fertil2)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.9519	-0.7184	0.0290	0.7384	7.3372

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.5913324	0.6450889	-5.567	2.74e-08	***
educ	-0.1639814	0.0655269	-2.503	0.0124	*
age	0.3281451	0.0190587	17.218	< 2e-16	***
age2	-0.0027222	0.0002766	-9.843	< 2e-16	***
electric	-0.1065314	0.1659650	-0.642	0.5210	
tv	-0.0025550	0.2092301	-0.012	0.9903	
bicycle	0.3320724	0.0515264	6.445	1.28e-10	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.479 on 4349 degrees of freedom

Multiple R-Squared: 0.5577, Adjusted R-squared: 0.5571

Wald test: 921.7 on 6 and 4349 DF, p-value: < 2.2e-16

- c) El modelo con variables instrumentales estima un coeficiente para educ de -0.164 , lo que significa que un año adicional de escolaridad reduce en promedio el número de hijos en 0.164 . Este efecto resulta ser el doble del obtenido mediante MCO, lo que indica que el modelo original estaba influenciado por problemas de endogeneidad y que, al no corregirlos, se subestimaba el verdadero impacto de la educación sobre la fertilidad.

El efecto más fuerte al corregir endogeneidad significa que la educación tiene un impacto aún mayor en reducir la cantidad de hijos de lo que se pensaba. En otras palabras, la educación es una política pública poderosa para influir en las tasas de natalidad.


```
> # -----
> # 6. Estimar modelo ampliado
> # -----
> modelo_14d_mco <- lm(children ~ educ + age + age2 + electric + tv + bicycle,
+                       data = fertil2)
> summary(modelo_14d_mco)

Call:
lm(formula = children ~ educ + age + age2 + electric + tv + bicycle,
    data = fertil2)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7781 -0.7205 -0.0169  0.7168  7.5645

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.3897837   0.2403173  -18.267  < 2e-16 ***
educ         -0.0767093   0.0063526  -12.075  < 2e-16 ***
age           0.3402038   0.0164417   20.692  < 2e-16 ***
age2        -0.0027081   0.0002706  -10.010  < 2e-16 ***
electric     -0.3027293   0.0761869   -3.974  7.20e-05 ***
tv           -0.2531443   0.0914374   -2.768  0.00566 **
bicycle       0.3178950   0.0493661    6.440  1.33e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.448 on 4349 degrees of freedom
(5 observations deleted due to missingness)
Multiple R-squared:  0.5761,    Adjusted R-squared:  0.5755
F-statistic: 984.9 on 6 and 4349 DF,  p-value: < 2.2e-16
```

- d) Al incorporar variables como acceso a electricidad, tenencia de televisión y bicicleta, el efecto estimado de la educación permanece negativo y con una magnitud similar a la obtenida mediante IV, lo que refuerza la evidencia de su influencia en la reducción de la fertilidad. Además, el hallazgo respecto a la televisión resulta particularmente relevante: contar con una está asociado, en promedio, con un cuarto de hijo menos, lo que podría relacionarse con una mayor exposición a información, cambios en las preferencias familiares o una redistribución del tiempo hacia actividades de ocio.

La influencia de la televisión refleja cómo los medios masivos transmiten información y modelos culturales que llevan a preferir familias más pequeñas, lo que muestra que la modernización y el acceso a bienes de consumo también afectan las decisiones reproductivas.

Ejercicio 15.

```
> # -----
> # 2. Estimar modelo por MCO
> # -----
> modelo_15a <- lm(Q ~ XPER + CAP + LAB, data = CHARD)
> summary(modelo_15a)
```

Call:

```
lm(formula = Q ~ XPER + CAP + LAB, data = CHARD)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-6.3447 -1.6842 -0.1289  1.3112  9.4533
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.76226	1.05535	1.670	0.099354	.
XPER	0.14684	0.06343	2.315	0.023517	*
CAP	0.43796	0.11756	3.725	0.000388	***
LAB	0.23916	0.09980	2.396	0.019195	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.756 on 71 degrees of freedom

Multiple R-squared: 0.5616, Adjusted R-squared: 0.543

F-statistic: 30.31 on 3 and 71 DF, p-value: 9.986e-13

- a) El modelo estimado mediante MCO evidencia que los tres factores —experiencia del administrador, capital y trabajo— aportan de manera positiva a la producción de vino. Entre ellos, el capital resulta ser el más determinante, pues una mayor inversión en este factor se asocia con incrementos significativos en la producción. El trabajo también muestra un efecto importante, seguido por la experiencia, que aunque de menor magnitud, mantiene significancia estadística. En conjunto, el modelo logra explicar alrededor del 56% de la variabilidad total de la producción, lo que representa un nivel de ajuste adecuado para este tipo de datos.

La inversión en capital (maquinaria, infraestructura) es la base para incrementar la producción, aunque la experiencia administrativa sigue siendo un activo clave. Esto refleja la combinación de recursos físicos y humanos en la eficiencia productiva.

```
> # -----  
> # 3.Intervalos de predicción para la producción de vino  
> # -----  
> # Calcular promedios de capital y trabajo  
> mean_cap <- mean(CHARD$CAP, na.rm = TRUE)  
> mean_lab <- mean(CHARD$LAB, na.rm = TRUE)  
>  
> # Crear nuevo data frame con los 3 escenarios  
> nuevos_datos <- data.frame(  
+   XPER = c(10, 20, 30),  
+   CAP = rep(mean_cap, 3),  
+   LAB = rep(mean_lab, 3)  
+ )  
>  
> # Predecir con intervalos de confianza al 95%  
> predicciones <- predict(modelo_15a, newdata = nuevos_datos, interval  
+   = "confidence")  
> predicciones  
      fit      lwr      upr  
1  9.064691 8.262487  9.866895  
2 10.533072 9.532133 11.534011  
3 12.001454 9.866095 14.136812  
>
```

- b) Aplicando el modelo MCO y manteniendo el capital y el trabajo en sus valores promedio, se estima que un administrador con 10 años de experiencia alcanza una producción de alrededor de 9.06 puntos en el índice. Con 20 años de experiencia, dicha producción aumenta a aproximadamente 10.53 puntos y con 30 años llega a 12 puntos. Esto confirma que, bajo condiciones similares de capital y trabajo, los administradores con mayor experiencia tienden a obtener una producción más elevada.

A mayor experiencia, mayor rendimiento, lo que confirma que la acumulación de conocimientos prácticos en la gestión de los viñedos se traduce directamente en mayor productividad.

```
> # -----
> # 4.Test de Hausman con residuos
> # -----
> reg_aux <- lm(XPER ~ AGE + CAP + LAB, data = CHARD)
> residuos_hausman <- resid(reg_aux)
>
> modelo_hausman <- lm(Q ~ XPER + CAP + LAB + residuos_hausman, data = CHARD)
> summary(modelo_hausman)
```

Call:

```
lm(formula = Q ~ XPER + CAP + LAB + residuos_hausman, data = CHARD)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-5.4582 -1.6581 -0.1773  1.4120  8.7115
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.48669	2.18960	-1.136	0.25996
XPER	0.51210	0.17731	2.888	0.00515 **
CAP	0.33213	0.12422	2.674	0.00933 **
LAB	0.23998	0.09721	2.469	0.01601 *
residuos_hausman	-0.41575	0.18917	-2.198	0.03127 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.685 on 70 degrees of freedom

Multiple R-squared: 0.5899, Adjusted R-squared: 0.5664

F-statistic: 25.17 on 4 and 70 DF, p-value: 6.14e-13

- c) La prueba de Hausman añade los residuos de una regresión auxiliar para evaluar si la experiencia está vinculada con factores omitidos en el modelo que influyen en la producción. El coeficiente de dichos residuos resulta significativo, lo que sugiere que la experiencia podría estar correlacionada con el término de error. En otras palabras, la estimación por MCO podría estar sesgada, lo que justifica la aplicación de un método alternativo que corrija este problema.

La endogeneidad de la experiencia implica que administradores más capaces tienden a permanecer más tiempo en el sector, sesgando los resultados. Económicamente, esto refleja que las habilidades gerenciales no observadas influyen tanto como los años de experiencia.

```
> # -----
> # 5.Estimación por VI
> # -----
> modelo_15d_vi <- ivreg(Q ~ XPER + CAP + LAB | AGE + CAP + LAB, data = CHARD)
> summary(modelo_15d_vi)

Call:
ivreg(formula = Q ~ XPER + CAP + LAB | AGE + CAP + LAB, data = CHARD)

Residuals:
    Min       1Q   Median       3Q      Max
-7.40413 -2.17750 -0.09044  2.28339 10.62769

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.4867     2.7226  -0.913   0.3642
XPER           0.5121     0.2205   2.323   0.0231 *
CAP            0.3321     0.1545   2.150   0.0349 *
LAB            0.2400     0.1209   1.985   0.0510 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.338 on 71 degrees of freedom
Multiple R-Squared:  0.3568,    Adjusted R-squared:  0.3296
Wald test: 21.24 on 3 and 71 DF, p-value: 6.307e-10
```

- d) Al estimar el modelo mediante variables instrumentales, utilizando la edad como instrumento de la experiencia, el efecto de esta sobre la producción se mantiene positivo y significativo, aunque con una magnitud algo mayor que la obtenida con MCO. Esto indica que, al corregir el posible sesgo, el impacto real de la experiencia en la producción de vino resulta más elevado de lo que sugería la estimación tradicional.

El efecto corregido más alto indica que la experiencia es aún más valiosa de lo que sugerían las estimaciones simples: con políticas de capacitación y retención de talento, la productividad del sector vitivinícola puede aumentar significativamente.

```
> # -----  
> # 6.Intervalos de predicción con modelo VI  
> # -----  
> mean_cap <- mean(CHARD$CAP, na.rm = TRUE)  
> mean_lab <- mean(CHARD$LAB, na.rm = TRUE)  
>  
> nuevos_datos_vi <- data.frame(  
+   XPER = c(10, 20, 30),  
+   CAP = rep(mean_cap, 3),  
+   LAB = rep(mean_lab, 3),  
+   AGE = c(35, 45, 55) # Se puede poner una edad razonable para mantener consistencia  
+ )  
>  
> predicciones_vi <- predict(modelo_15d_vi, newdata = nuevos_datos_vi,  
+                           interval = "confidence")  
> predicciones_vi  
      1      2      3  
7.647467 12.768487 17.889508  
>
```

e) Con el modelo de variables instrumentales, y manteniendo el capital y el trabajo en sus valores promedio, se estiman producciones de 7.65 puntos para administradores con 10 años de experiencia, 12.77 para aquellos con 20 años y 17.89 para quienes tienen 30 años. Esto refleja un aumento más marcado en la producción a medida que crece la experiencia, lo que refuerza la importancia de este factor como elemento clave para optimizar los resultados en los viñedos.

El crecimiento de la producción con la experiencia muestra la importancia de la estabilidad y profesionalización en la gestión agrícola: mantener administradores experimentados es tan importante como invertir en maquinaria.

Bono:

```

> # =====
> # 2.5 Unir ingreso + gasto y crear gasto per cápita
> # =====
> base_final <- ing_hogar |>
+   inner_join(gas_hogar, by = c("vivienda","hogar")) |>
+   mutate(gasto_pc = gasto_total / num_miembros)
>
> # chequeo rápido
> glimpse(base_final)
Rows: 8,892
Columns: 12
$ vivienda      <dbl> 1, 3, 4, 6, 7, 9, 10, 11, 13, 14, 15, 17, 18, 19, 21, ...
$ hogar         <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
$ ingreso_total <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
$ num_miembros  <int> 3, 1, 3, 3, 3, 2, 2, 3, 3, 2, 1, 3, 3, 3, 2, 2, 2, ...
$ quintil       <dbl> 1, 1, 5, 2, 3, 4, 3, 4, 3, 4, 2, 1, 2, 2, 1, 4, 4, ...
$ des_provincia <chr> "DISTRITO NACIONAL", "DISTRITO NACIONAL", "DISTRITO NA...
$ escolaridad   <dbl> 7, 5, 12, 12, 11, 13, 11, 15, 12, 8, 8, 1, 8, 8, 12, 2...
$ grupo_edad    <chr> "25 - 39", "60 y más", "25 - 39", "25 - 39", "25 - 39"...
$ grupo_educacion <chr> "Primario", "Primario", "Secundario", "Secundario", "S...
$ ingreso_pc    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
$ gasto_total   <dbl> 14314.00, 2847.00, 82260.41, 25425.26, 24600.26, 29298...
$ gasto_pc      <dbl> 4771.333, 2847.000, 27420.137, 8475.087, 8200.087, 146...
> summary(base_final[, c("ingreso_pc","gasto_pc","num_miembros")])
  ingreso_pc      gasto_pc      num_miembros
Min.   : 0.0    Min.   : 0    Min.   : 1.000
1st Qu.: 0.0    1st Qu.: 4038  1st Qu.: 2.000
Median : 0.0    Median : 6594  Median : 3.000
Mean   : 266.2  Mean   : 8758  Mean   : 3.193
3rd Qu.: 0.0    3rd Qu.: 10560  3rd Qu.: 4.000
Max.   :48552.9  Max.   :167196  Max.   :13.000
> head(base_final)
> head(base_final)
# A tibble: 6 × 12
  vivienda hogar ingreso_total num_miembros quintil des_provincia escolaridad
  <dbl> <dbl>      <dbl>      <int>    <dbl> <chr>      <dbl>
1      1      1          0          3      1 DISTRITO NACION...      7
2      3      1          0          1      1 DISTRITO NACION...      5
3      4      1          0          3      5 DISTRITO NACION...     12
4      6      1          0          3      2 DISTRITO NACION...     12
5      7      1          0          3      3 DISTRITO NACION...     11
6      9      1          0          2      4 DISTRITO NACION...     13
# i 5 more variables: grupo_edad <chr>, grupo_educacion <chr>,
#   ingreso_pc <dbl>, gasto_total <dbl>, gasto_pc <dbl>

```

```
> # =====
> # 3.2 Modelo log-log (elasticidad)
> # E[gasto_pc | ingreso_pc, controles]
> # =====
> mod <- lm(log(gasto_pc) ~ log(ingreso_pc) + escolaridad +
+           quintil + grupo_edad + des_provincia, data = df)
> summary(mod)
```

Call:

```
lm(formula = log(gasto_pc) ~ log(ingreso_pc) + escolaridad +
    quintil + grupo_edad + des_provincia, data = df)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.38459	-0.35668	0.03735	0.31676	1.55737

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.50648	0.50341	12.925	< 2e-16	***
log(ingreso_pc)	0.11050	0.05004	2.208	0.02826	*
escolaridad	0.01679	0.01033	1.627	0.10525	
quintil2	0.17573	0.17470	1.006	0.31556	
quintil3	0.42892	0.17778	2.413	0.01666	*
quintil4	0.39884	0.19453	2.050	0.04152	*
quintil5	1.17662	0.21191	5.552	8.08e-08	***
grupo_edad25 - 39	-0.02733	0.24623	-0.111	0.91171	
grupo_edad40 - 59	0.06161	0.24077	0.256	0.79827	
grupo_edad60 y más	0.11678	0.25301	0.462	0.64485	
des_provinciaBARAHONA	-0.05203	0.68294	-0.076	0.93934	
des_provinciaDISTRITO NACIONAL	1.04075	0.26969	3.859	0.00015	***
des_provinciaDUARTE	2.07588	0.68142	3.046	0.00260	**
des_provinciaEL SEIBO	0.47363	0.36641	1.293	0.19751	
des_provinciaESPAILLAT	0.67330	0.30657	2.196	0.02912	*

des_provinciaDUARTE	2.07588	0.68142	3.046	0.00260	**
des_provinciaEL SEIBO	0.47363	0.36641	1.293	0.19751	
des_provinciaESPAILLAT	0.67330	0.30657	2.196	0.02912	*
des_provinciaHATO MAYOR	0.54362	0.39619	1.372	0.17142	
des_provinciaINDEPENDENCIA	0.64278	0.50048	1.284	0.20038	
des_provinciaLA ALTAGRACIA	0.95243	0.34544	2.757	0.00632	**
des_provinciaLA ROMANA	0.63894	0.29744	2.148	0.03280	*
des_provinciaLA VEGA	1.09341	0.32422	3.372	0.00088	***
des_provinciaMARIA TRINIDAD SANCHEZ	1.35980	0.49801	2.730	0.00684	**
des_provinciaMONSEÑOR NOUEL	0.47454	0.32889	1.443	0.15049	
des_provinciaMONTE CRISTI	1.11961	0.66895	1.674	0.09561	.
des_provinciaMONTE PLATA	0.43655	0.36163	1.207	0.22865	
des_provinciaPEDERNALES	0.32154	0.67385	0.477	0.63371	
des_provinciaPERAVIA	0.37722	0.42519	0.887	0.37595	
des_provinciaPUERTO PLATA	1.03282	0.35601	2.901	0.00410	**
des_provinciasALCEDO	1.07913	0.43424	2.485	0.01370	*
des_provinciasSAN CRISTOBAL	0.68000	0.27957	2.432	0.01580	*
des_provinciasSAN JOSE DE OCOA	1.10658	0.49656	2.228	0.02686	*
des_provinciasSAN JUAN	-0.12000	0.27056	-0.444	0.65782	
des_provinciasSAN PEDRO DE MACORIS	0.01230	0.29498	0.042	0.96677	
des_provinciasSANCHEZ RAMIREZ	0.93566	0.43314	2.160	0.03184	*
des_provinciasSANTIAGO	0.78310	0.26372	2.969	0.00331	**
des_provinciasSANTIAGO RODRIGUEZ	1.09204	0.50575	2.159	0.03191	*
des_provinciasSANTO DOMINGO	0.64880	0.25009	2.594	0.01012	*
des_provinciaVALVERDE	0.67386	0.36423	1.850	0.06564	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6173 on 220 degrees of freedom

Multiple R-squared: 0.6404, Adjusted R-squared: 0.5816

F-statistic: 10.88 on 36 and 220 DF, p-value: < 2.2e-16

- a) El modelo log-log estimado muestra que la elasticidad del gasto per cápita respecto al ingreso per cápita es cercana a 0.111. Esto implica que, manteniendo constantes las demás variables, un aumento del 1% en el ingreso per cápita se relaciona con un incremento del 0.111% en el gasto per cápita. El coeficiente resulta estadísticamente significativo al 5%, lo que respalda la solidez de la relación estimada desde el punto de vista estadístico.

La elasticidad menor a 1 indica que, aunque el gasto aumenta con el ingreso, lo hace en menor proporción. Esto refleja que los hogares satisfacen primero necesidades básicas antes de destinar ingreso extra a otros consumos.

```
>
> # Elasticidad ingreso del gasto = coef de log(ingreso_pc)
> elasticidad <- coef(mod)["log(ingreso_pc)"]; elasticidad
log(ingreso_pc)
0.1105046
>
> # =====
> # 3.3 Heterocedasticidad
> # =====
> # Breusch-Pagan
> bp <- bptest(mod) ; bp

studentized Breusch-Pagan test

data: mod
BP = 39.72, df = 36, p-value = 0.3078
> # white
> white <- bptest(mod, ~ fitted(mod) + I(fitted(mod)^2)) ; white

studentized Breusch-Pagan test

data: mod
BP = 11.31, df = 2, p-value = 0.0035
```

- b) La elasticidad estimada evidencia que el gasto per cápita responde de forma inelástica a las variaciones en el ingreso per cápita. Si bien en este resultado no se diferencian modelos por grupos específicos, la baja magnitud sugiere que, en el agregado de los hogares, el gasto per cápita varía poco en relación al ingreso, lo cual es coherente con el comportamiento de los bienes de primera necesidad.

La inelasticidad es coherente con la naturaleza de bienes esenciales como alimentación, vivienda o servicios básicos, donde el consumo no aumenta demasiado incluso si suben los ingresos.

- c) La prueba de Breusch-Pagan arroja un valor p de 0.308, por lo que no se rechaza la hipótesis nula de homocedasticidad en este caso. En contraste, la prueba de White obtiene un valor p de 0.0035, lo que sí proporciona evidencia de heterocedasticidad. Esta discrepancia se explica porque la prueba de White es capaz de detectar patrones más generales, no necesariamente lineales, lo que la hace más sensible en determinadas situaciones.

La heterocedasticidad encontrada con la prueba de White refleja que el impacto del ingreso sobre el gasto no es igual en todos los hogares: las familias más pobres y más ricas responden de forma distinta a incrementos en el ingreso.

```
> # =====
> # 3.4 Errores estándar robustos
> # =====
> rob <- coeftest(mod, vcov = vcovHC(mod, type = "HC1"))
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.506476	0.429794	15.1386	< 2.2e-16	***
log(ingreso_pc)	0.110505	0.049919	2.2137	0.0278777	*
escolaridad	0.016795	0.012091	1.3890	0.1662277	
quintil2	0.175733	0.188107	0.9342	0.3512160	
quintil3	0.428918	0.187132	2.2921	0.0228478	*
quintil4	0.398845	0.204121	1.9540	0.0519727	.
quintil5	1.176620	0.224601	5.2387	3.778e-07	***
grupo_edad25 - 39	-0.027334	0.167742	-0.1630	0.8707043	
grupo_edad40 - 59	0.061611	0.171189	0.3599	0.7192674	
grupo_edad60 y más	0.116780	0.208651	0.5597	0.5762611	
des_provinciaBARAHONA	-0.052034	0.267756	-0.1943	0.8460929	
des_provinciaDISTRITO NACIONAL	1.040755	0.229379	4.5373	9.371e-06	***
des_provinciaDUARTE	2.075882	0.258254	8.0381	5.552e-14	***
des_provinciaEL SEIBO	0.473626	0.218636	2.1663	0.0313649	*
des_provinciaESPAILLAT	0.673304	0.217719	3.0925	0.0022414	**
des_provinciaHATO MAYOR	0.543622	0.316343	1.7185	0.0871204	.
des_provinciaINDEPENDENCIA	0.642781	0.196237	3.2755	0.0012251	**
des_provinciaLA ALTAGRACIA	0.952430	0.222749	4.2758	2.839e-05	***
des_provinciaLA ROMANA	0.638939	0.250960	2.5460	0.0115810	*
des_provinciaLA VEGA	1.093410	0.241626	4.5252	9.874e-06	***
des_provinciaMARIA TRINIDAD SANCHEZ	1.359795	0.566167	2.4018	0.0171476	*
des_provinciaONSEÑOR NOEL	0.474537	0.261380	1.8155	0.0708080	.
des_provinciaMONTE CRISTI	1.119607	0.213904	5.2342	3.862e-07	***
des_provinciaMONTE PLATA	0.436554	0.335167	1.3025	0.1941080	
des_provinciaPEDERNALES	0.321543	0.237573	1.3535	0.1773008	
des_provinciaPERAVIA	0.377224	0.400209	0.9426	0.3469357	
des_provinciaPUERTO PLATA	1.032819	0.258094	4.0017	8.594e-05	***
des_provinciaSALCEDO	1.079128	0.311370	3.4657	0.0006357	***
des_provinciaSAN CRISTOBAL	0.679998	0.231504	2.9373	0.0036630	**
des_provinciaSAN JOSE DE OCOA	1.106581	0.249392	4.4371	1.441e-05	***
des_provinciaSAN JUAN	-0.120000	0.277410	-0.4326	0.6657488	

```

des_provinciaBARAHONA      -0.052034    0.267756  -0.1943  0.8460929
des_provinciaDISTRITO NACIONAL  1.040755    0.229379   4.5373  9.371e-06 ***
des_provinciaDUARTE        2.075882    0.258254   8.0381  5.552e-14 ***
des_provinciaEL SEIBO      0.473626    0.218636   2.1663  0.0313649 *
des_provinciaESPAILLAT     0.673304    0.217719   3.0925  0.0022414 **
des_provinciaHATO MAYOR    0.543622    0.316343   1.7185  0.0871204 .
des_provinciaINDEPENDENCIA 0.642781    0.196237   3.2755  0.0012251 **
des_provinciaLA ALTAGRACIA 0.952430    0.222749   4.2758  2.839e-05 ***
des_provinciaLA ROMANA     0.638939    0.250960   2.5460  0.0115810 *
des_provinciaLA VEGA       1.093410    0.241626   4.5252  9.874e-06 ***
des_provinciaMARIA TRINIDAD SANCHEZ 1.359795    0.566167   2.4018  0.0171476 *
des_provinciaMONSEÑOR NOUEL 0.474537    0.261380   1.8155  0.0708080 .
des_provinciaMONTE CRISTI  1.119607    0.213904   5.2342  3.862e-07 ***
des_provinciaMONTE PLATA   0.436554    0.335167   1.3025  0.1941080
des_provinciaPEDERNALES    0.321543    0.237573   1.3535  0.1773008
des_provinciaPERAVIA       0.377224    0.400209   0.9426  0.3469357
des_provinciaPUERTO PLATA  1.032819    0.258094   4.0017  8.594e-05 ***
des_provinciaSALCEDO       1.079128    0.311370   3.4657  0.0006357 ***
des_provinciasAN CRISTOBAL 0.679998    0.231504   2.9373  0.0036630 **
des_provinciasAN JOSE DE OCOA 1.106581    0.249392   4.4371  1.441e-05 ***
des_provinciasAN JUAN      -0.120000    0.277410  -0.4326  0.6657488
des_provinciasAN PEDRO DE MACORIS 0.012301    0.288214   0.0427  0.9659948
des_provinciasANCHEZ RAMIREZ 0.935664    0.213675   4.3789  1.845e-05 ***
des_provinciasANTIAGO      0.783101    0.213445   3.6689  0.0003057 ***
des_provinciasANTIAGO RODRIGUEZ 1.092042    0.269758   4.0482  7.150e-05 ***
des_provinciasANTO DOMINGO 0.648798    0.185393   3.4996  0.0005639 ***
des_provinciaAVALVERDE     0.673860    0.263247   2.5598  0.0111438 *
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

d) Dada la evidencia de heterocedasticidad identificada mediante la prueba de White, el modelo se reestimó empleando errores estándar robustos tipo HC1. Esta corrección permite ajustar las inferencias y mitigar posibles sesgos en los errores estándar y, en consecuencia, en las pruebas de significancia. Con el uso de errores robustos, los coeficientes y sus signos se mantienen inalterados, aunque algunos valores p se modifican, lo que confirma la significancia estadística de variables como el ingreso per cápita, varios quintiles de ingreso y determinadas provincias.

Usar errores robustos asegura que se pueda evaluar correctamente la influencia del ingreso en diferentes quintiles y regiones, lo que es crucial para diseñar políticas sociales.

```

> # =====
> # 3.5 Resumen
> # =====
> cat("\nElasticidad ingreso del gasto (log-log): ", round(elasticidad, 3), "\n")

Elasticidad ingreso del gasto (log-log): 0.111
> cat("BP p-value: ", signif(bp$p.value, 3),
+      " | white p-value: ", signif(white$p.value, 3), "\n")
BP p-value: 0.308 | white p-value: 0.0035
> cat("Use esta tabla con errores robustos para inferencia:\n")
Use esta tabla con errores robustos para inferencia:
> print(rob)

t test of coefficients:


```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.506476	0.429794	15.1386	< 2.2e-16	***
log(ingreso_pc)	0.110505	0.049919	2.2137	0.0278777	*
escolaridad	0.016795	0.012091	1.3890	0.1662277	
quintil2	0.175733	0.188107	0.9342	0.3512160	
quintil3	0.428918	0.187132	2.2921	0.0228478	*
quintil4	0.398845	0.204121	1.9540	0.0519727	.
quintil5	1.176620	0.224601	5.2387	3.778e-07	***
grupo_edad25 - 39	-0.027334	0.167742	-0.1630	0.8707043	
grupo_edad40 - 59	0.061611	0.171189	0.3599	0.7192674	
grupo_edad60 y más	0.116780	0.208651	0.5597	0.5762611	
des_provinciaABARAHONA	-0.052034	0.267756	-0.1943	0.8460929	
des_provinciaDISTRITO NACIONAL	1.040755	0.229379	4.5373	9.371e-06	***
des_provinciaDUARTE	2.075882	0.258254	8.0381	5.552e-14	***
des_provinciaEL SEIBO	0.473626	0.218636	2.1663	0.0313649	*
des_provinciaESPAILLAT	0.673304	0.217719	3.0925	0.0022414	**
des_provinciaHATO MAYOR	0.543622	0.316343	1.7185	0.0871204	.
grupo_edad25 - 39	-0.027334	0.167742	-0.1630	0.8707043	
grupo_edad40 - 59	0.061611	0.171189	0.3599	0.7192674	
grupo_edad60 y más	0.116780	0.208651	0.5597	0.5762611	
des_provinciaABARAHONA	-0.052034	0.267756	-0.1943	0.8460929	
des_provinciaDISTRITO NACIONAL	1.040755	0.229379	4.5373	9.371e-06	***
des_provinciaDUARTE	2.075882	0.258254	8.0381	5.552e-14	***
des_provinciaEL SEIBO	0.473626	0.218636	2.1663	0.0313649	*
des_provinciaESPAILLAT	0.673304	0.217719	3.0925	0.0022414	**
des_provinciaHATO MAYOR	0.543622	0.316343	1.7185	0.0871204	.
des_provinciaINDEPENDENCIA	0.642781	0.196237	3.2755	0.0012251	**
des_provinciaLA ALTAGRACIA	0.952430	0.222749	4.2758	2.839e-05	***
des_provinciaLA ROMANA	0.638939	0.250960	2.5460	0.0115810	*
des_provinciaLA VEGA	1.093410	0.241626	4.5252	9.874e-06	***
des_provinciaMARIA TRINIDAD SANCHEZ	1.359795	0.566167	2.4018	0.0171476	*
des_provinciaMONSEÑOR NOUEL	0.474537	0.261380	1.8155	0.0708080	.
des_provinciaMONTE CRISTI	1.119607	0.213904	5.2342	3.862e-07	***
des_provinciaMONTE PLATA	0.436554	0.335167	1.3025	0.1941080	
des_provinciaPEDERNALES	0.321543	0.237573	1.3535	0.1773008	
des_provinciaPERAVIA	0.377224	0.400209	0.9426	0.3469357	
des_provinciaPUERTO PLATA	1.032819	0.258094	4.0017	8.594e-05	***
des_provinciaSALCEDO	1.079128	0.311370	3.4657	0.0006357	***
des_provinciaSAN CRISTOBAL	0.679998	0.231504	2.9373	0.0036630	**
des_provinciaSAN JOSE DE OCOA	1.106581	0.249392	4.4371	1.441e-05	***
des_provinciaSAN JUAN	-0.120000	0.277410	-0.4326	0.6657488	
des_provinciaSAN PEDRO DE MACORIS	0.012301	0.288214	0.0427	0.9659948	
des_provinciaSANCHEZ RAMIREZ	0.935664	0.213675	4.3789	1.845e-05	***
des_provinciaSANTIAGO	0.783101	0.213445	3.6689	0.0003057	***
des_provinciaSANTIAGO RODRIGUEZ	1.092042	0.269758	4.0482	7.150e-05	***
des_provinciaSANTO DOMINGO	0.648798	0.185393	3.4996	0.0005639	***
des_provinciaVALVERDE	0.673860	0.263247	2.5598	0.0111438	*

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

e) Es posible que algunas variables explicativas, como el ingreso per cápita, presenten endogeneidad debido a problemas de causalidad inversa o a la omisión de variables relevantes, lo que podría sesgar las estimaciones obtenidas por MCO. Una alternativa para corregir este problema es emplear variables instrumentales que estén correlacionadas con el ingreso, pero no con el término de error, o bien aplicar un enfoque de función de control. Como ejemplos de instrumentos potenciales se podrían considerar características geográficas o históricas que influyan en el ingreso, pero no afecten directamente al gasto.

La posible endogeneidad del ingreso refleja que no es sólo el ingreso el que determina el gasto: también el gasto (en educación, salud, etc.) puede influir en el ingreso futuro. Por eso se necesitan instrumentos que permitan identificar el efecto puro del ingreso.