

# Examen #2

## Métodos Cuantitativos Aplicados, CIDE A.C.

### Instrucciones

El examen contiene una sección con 11 preguntas sobre regresión lineal (75 puntos en total). Tienes 90 minutos para contestarlo. Te recomiendo leer las preguntas con cuidado y tratar de no dejar respuestas en blanco. Si no tienes la respuesta final, siempre es mejor mostrar el procedimiento que tu crees que te llevaría a dicha respuesta.

Al terminar, deberás enviar el archivo con tus respuestas a mi dirección de correo electrónico ([sergio.bejar@cide.edu](mailto:sergio.bejar@cide.edu)). ¡Buena suerte!

### Sección 1: Actitudes sobre Gasto de Gobierno en Bienes Públicos

Las preguntas en esta sección serán respondidas con la base de datos `gss.csv` disponible en la carpeta “Files” de la página Github de nuestra clase. Los datos vienen de la primera ola de la encuesta General Social Survey en los Estados Unidos. En ella, se tratan de investigar las actitudes de los ciudadanos con respecto al gasto nacional en bienes públicos.

Las variables de nuestro interés son las siguientes:

**educ:** educación, mide el número de años de escuela completados (min: 0; max: 20).

**race:** raza del individuo (1 = blanco, 2 = negro, 3 = otra).

**age:** edad (min: 18, max: 89).

**polviews:** ideología (min: 1-extremadamente liberal-, max: 7-extremadamente conservador-).

**sumnatsoc:** suma de opiniones respecto a gasto en varios bienes públicos (i.e. `natenvir`, `natheal`, `natdrug`, `nateduc`, `natrace`, `natfare`, `natroad`, `natmass`, `natpark`, `natsoc`, `natchld`). Es una variable de intervalo con una distribución casi normal. Valores más altos indican que el individuo tiene opiniones más favorables respecto a incrementar el gasto de gobierno en bienes públicos.

Recodifica `race` para tener una dummy que tome el valor de 1 para individuos de raza negra y 0 para otras razas. Denota esta variable como `raza_d`. (3 pts.)

1.

```
gss <- gss %>% mutate(raza_d = if_else(race=="2", 1, 0))
```

Estima un modelo OLS que tenga la siguiente especificación: (5 pts.)

1.

$$\text{sumnatsoc} = f(\text{educacín}, \text{raza\_d}, \text{edad}, \text{ideología})$$

```
m_e_1 <- lm(sumnatsoc ~ educ + age + raza_d + polviews, data = gss) ## estimo modelo
```

Presenta los resultados en una tabla bien legible (i.e. no usando `summary()` sino un paquete que nos permita presentar mejor los resultados. (5 pts.)

```
stargazer(m_e_1,
  covariate.labels = c("educación", "edad", "raza = afro-americana",
  dep.var.labels = c("Opinión sobre Gasto en Bienes Públicos"),
  header=FALSE,
  type='text') ## Uso paquete `stargazer`
```

Dependent variable:	
Opinión sobre Gasto en Bienes Públicos	
educación	0.077*** (0.023)
edad	-0.002 (0.004)
raza = afro-americana	1.592*** (0.185)
ideología	-0.752*** (0.045)
intercepto	7.102***

(0.423)

```
-----
Observations                2,239
R2                          0.154
Adjusted R2                 0.153
Residual Std. Error        3.155 (df = 2234)
F Statistic                101.768*** (df = 4; 2234)
=====
Note:                       *p<0.1; **p<0.05; ***p<0.01
```

Explica los resultados poniendo énfasis especial en los coeficientes de las variables de educación, raza (la dummy que creaste), e ideología. (10 pts.)

**RESPUESTA:** En este modelo hay 3 variables que son estadísticamente significativas al 95% de confianza. Dichas variables son: **edad** **raza** = **afro-americana** e **ideología**. Substantivamente los efectos de dichas variables son los siguientes:

- La **edad** tiene un efecto positivo sobre la variable dependiente. Específicamente, por cada aumento de una unidad en dicha variable, la variable dependiente (**Opinión sobre Gasto en Bienes Públicos**) se incrementa en 1.59 unidades.
- También hay un efecto positivo de la variable **ideología**. Por cada unidad de incremento en esta variable, la variable dependiente se incrementa en 7.1 unidades.
- La variable dummy de **raza** tiene un efecto negativo sobre la variable dependiente. Específicamente, la opinión que tienen los afro-americanos sobre el gasto en bienes públicos es .7 unidades más baja que la opinión de las otras razas.
- No hay efecto significativo estadístico entre la variable de **educación** y la variable dependiente **Opinión sobre Gasto en Bienes Públicos**).

Estima las pruebas de Durbin-Watson y Breush-Godfrey y explica detalladamente por que sí o por que no existe un problema de autocorrelación en los datos. (10 pts.)

```
dwtest(m_e_1) ## funcion para calcular prueba D-W
```

Durbin-Watson test

```
data:  m_e_1
DW = 1.9847, p-value = 0.3538
alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(m_e_1) ## funcion para calbular prueba B-G
```

Breusch-Godfrey test for serial correlation of order up to 1

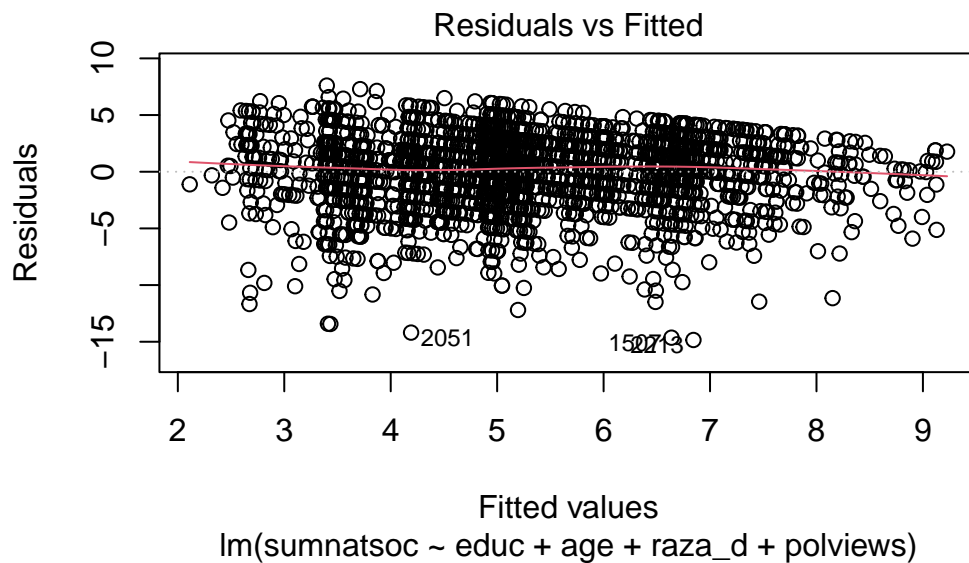
```
data: m_e_1
```

```
LM test = 0.13, df = 1, p-value = 0.7184
```

- El  $p$ -value de la prueba D-W es igual a 0.3538 y el de la prueba B-G es 0.7184. En ambos casos los valores son mayores a 0.05 por lo que se acepta la hipótesis nula de no-autocorrelación.

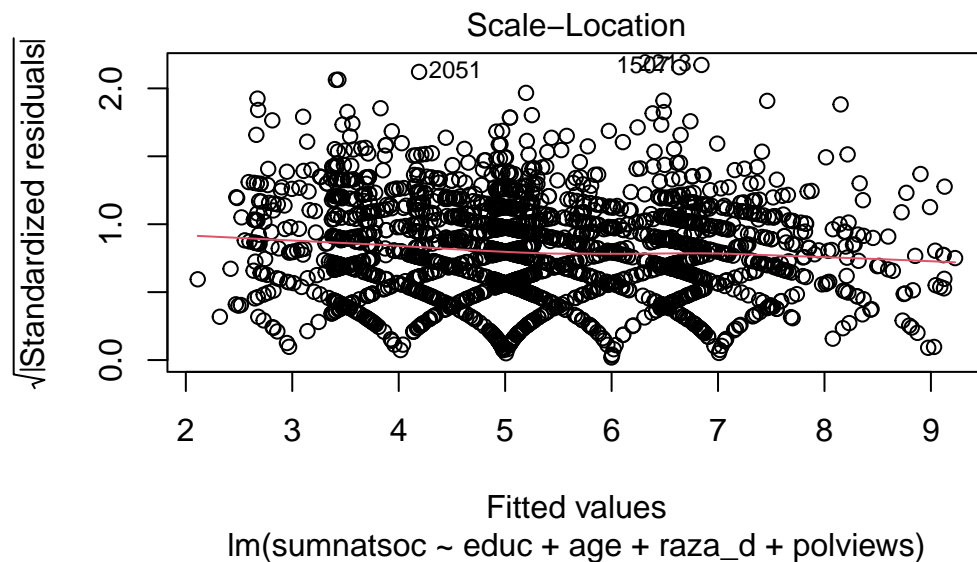
Checa la linealidad del modelo que estimaste en (2) y determina hay problemas de linealidad. Explica detalladamente tu respuesta. (7 pts.)

```
plot(m_e_1, which = 1)
```



Elabora dos pruebas de diagnóstico diferentes para determinar si existe problema de heteroskedasticidad en la regresión que estimaste. Explica cuidadosamente tu respuesta. (10 pts.)

```
plot(m_e_1, which = 3) ## gráfico para ver dispersión de puntos
```



```
bptest(m_e_1) ## prueba Breusch-Pagan
```

studentized Breusch-Pagan test

data: m\_e\_1

BP = 24.386, df = 4, p-value = 6.684e-05

En el gráfico hay patrones claros en la distribución de los puntos y en la prueba Breusch-Pagan el *p-value* es menor a 0.05 por lo que hay evidencia de heteroskedasticidad.

Estima de nuevo el modelo pero ahora con errores estándar robustos (función `coeftest` de paquete “sandwich”). ¿Qué diferencias encuentras en los resultados? (10 pts.)

```
library(sandwich)
coeftest(m_e_1, vcov = vcovHC)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.1015484	0.4693482	15.1307	< 2.2e-16 ***
educ	0.0765046	0.0247343	3.0930	0.002006 **
age	-0.0015243	0.0038427	-0.3967	0.691650
raza_d	1.5919757	0.1770660	8.9909	< 2.2e-16 ***
polviews	-0.7522213	0.0496337	-15.1554	< 2.2e-16 ***

---

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

Una forma de minimizar problemas de heteroskedasticidad es estimando una regresión con errores estándar robustos, que es lo que estamos haciendo en este caso. Hay ligeros cambios en algunos coeficientes estimados y en el error estándar de los coeficientes en la nueva regresión, aunque no hay cambios significativos en las inferencias que obtuvimos en el modelo que estimamos primero.

El profe piensa que existe la posibilidad de que el efecto de ideología política sea condicional dependiendo de la raza de los individuos. Estima un modelo que evalúe la hipótesis del profe. (5 pts.)

```
m_e_2 <- lm(sumnatsoc ~ educ + age + raza_d + polviews + polviews*raza_d, data = gss)
summary(m_e_2) ## summary del modelo con interaccion
```

Call:

```
lm(formula = sumnatsoc ~ educ + age + raza_d + polviews + polviews *
    raza_d, data = gss)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.0819	-1.8167	0.3937	2.1122	7.7512

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.379600	0.426331	17.310	< 2e-16 ***
educ	0.079101	0.022971	3.444	0.000585 ***
age	-0.001284	0.003741	-0.343	0.731467
raza_d	-0.524261	0.516108	-1.016	0.309836
polviews	-0.831684	0.048442	-17.168	< 2e-16 ***
raza_d:polviews	0.558524	0.127266	4.389	1.19e-05 ***

---

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

Residual standard error: 3.142 on 2233 degrees of freedom

(109 observations deleted due to missingness)

Multiple R-squared: 0.1614, Adjusted R-squared: 0.1595

F-statistic: 85.93 on 5 and 2233 DF, p-value: < 2.2e-16

Elabora un gráfico que muestre el efecto condicional de ideología política en la variable dependiente. (5 pts.)

```
library(interplot)
interplot(m_e_2,
          var1 = "polviews",
          var2 = "raza_d")+
  labs(x = "dummy de raza",
       y = "Efecto Condicional de ideología")
```



Explica detalladamente los resultados del gráfico que elaboraste en el inciso anterior. (5 pts.)

Hay mediación de la raza en el efecto negativo de la ideología. Dicho efecto es más negativo para individuos de raza diferente a afro-americanos ( $raza\_d = 0$ ) que para afro-americanos ( $raza\_d = 1$ ).