

Examen I

Econometría Básica 2017-II
Paul Rodríguez Lesmes

11/09/2017

Indicaciones

- Tiene 1 hora y 50 minutos para responder todo el examen
- Para poder hacer reclamos, responda este examen con tinta permanente
- Debe responder en los espacios indicados, no se aceptan hojas suplementarias
- Siga exactamente la notación que se encuentra en su enunciado. De lo contrario se penalizará con el 50 % de la nota de cada punto en la que se presente
- El examen consta de 50 puntos, distribuidos en las preguntas según se expresa en cada pregunta.

1. Parte I: Conceptos (10 puntos)

Pregunta 1 (2pt). El sesgo de selección, $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ dado que D_i indica si el individuo recibió el tratamiento o no, se elimina cuando el efecto causal es el mismo para todos los integrantes de una población. Es decir, si todos nos beneficiamos de exactamente la misma manera, de una intervención.

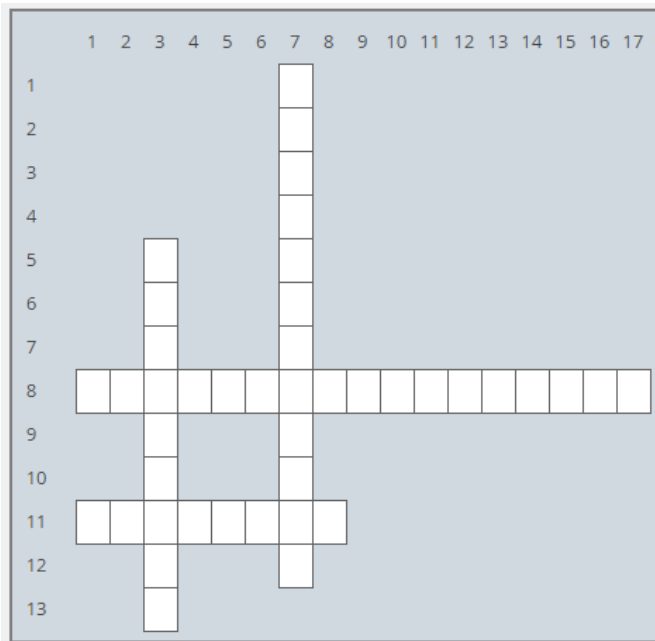
- a. Verdadero
- b. Falso

Pregunta 2 (2pt). Si $F(x)$ es una CDF válida sobre una variable continua X , y tenemos que $x_1 > x_2$, entonces $F(x_1) > F(x_2)$

- a. Verdadero
- b. Falso

Pregunta 3 (2pt). El modelo $y_i = \beta_1 + \beta_2 \left(\frac{1}{x_i}\right) + u_i$ es lineal en parámetros:

- a. Verdadero
- b. Falso



Horizontal

8: Si x y u son dos variable aleatorias, y es el caso que $Var(u|x) = \sigma^2$, es decir, es constante, decimos que se cumple esta propiedad

11: El cuarto momento central de una distribución. Es una medida del grosor de las colas de la misma.

Vertical

3: El tercer momento central de una distribución. Nos da una medida de qué tan parecida es la parte derecha e izquierda de una distribución entorno a su media.

7: Dado un estimador $\hat{\theta}$, se cumple esta propiedad si $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$, y $\lim_{n \rightarrow \infty} Var(\hat{\theta}) = 0$

2. Parte II: Teoría Econométrica (25 puntos)

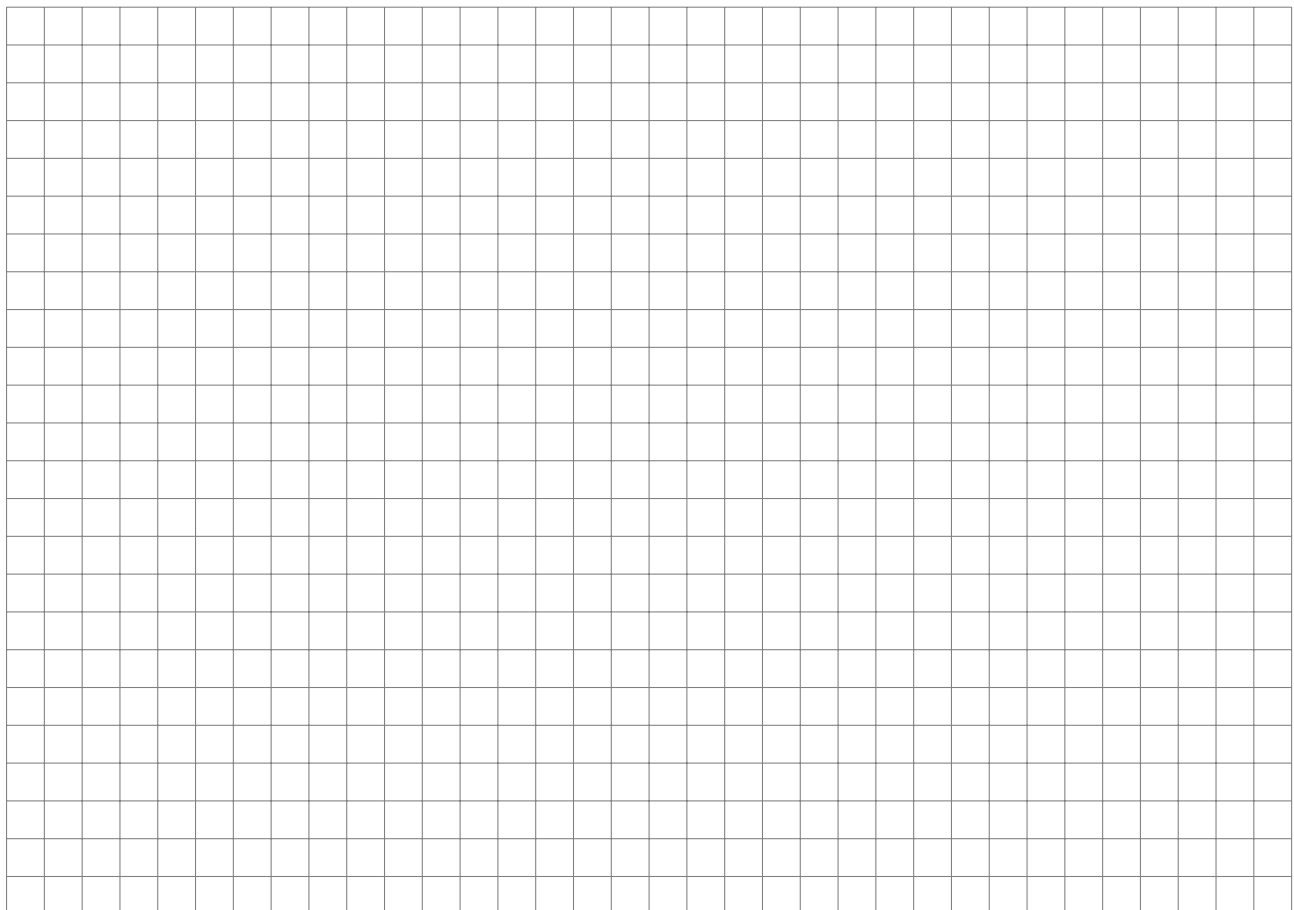
Dado el siguiente modelo poblacional

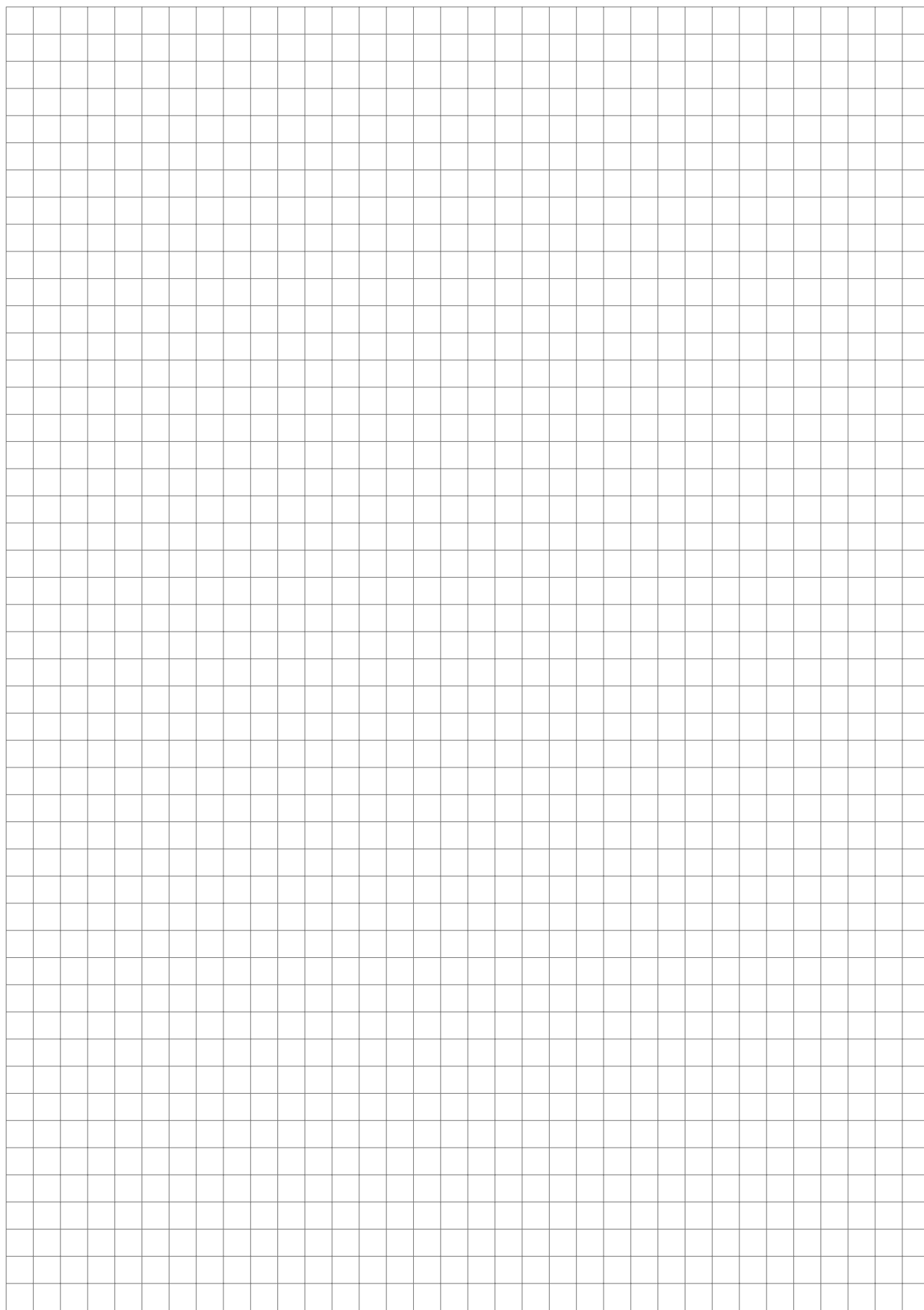
$$Y = Z\theta + u$$

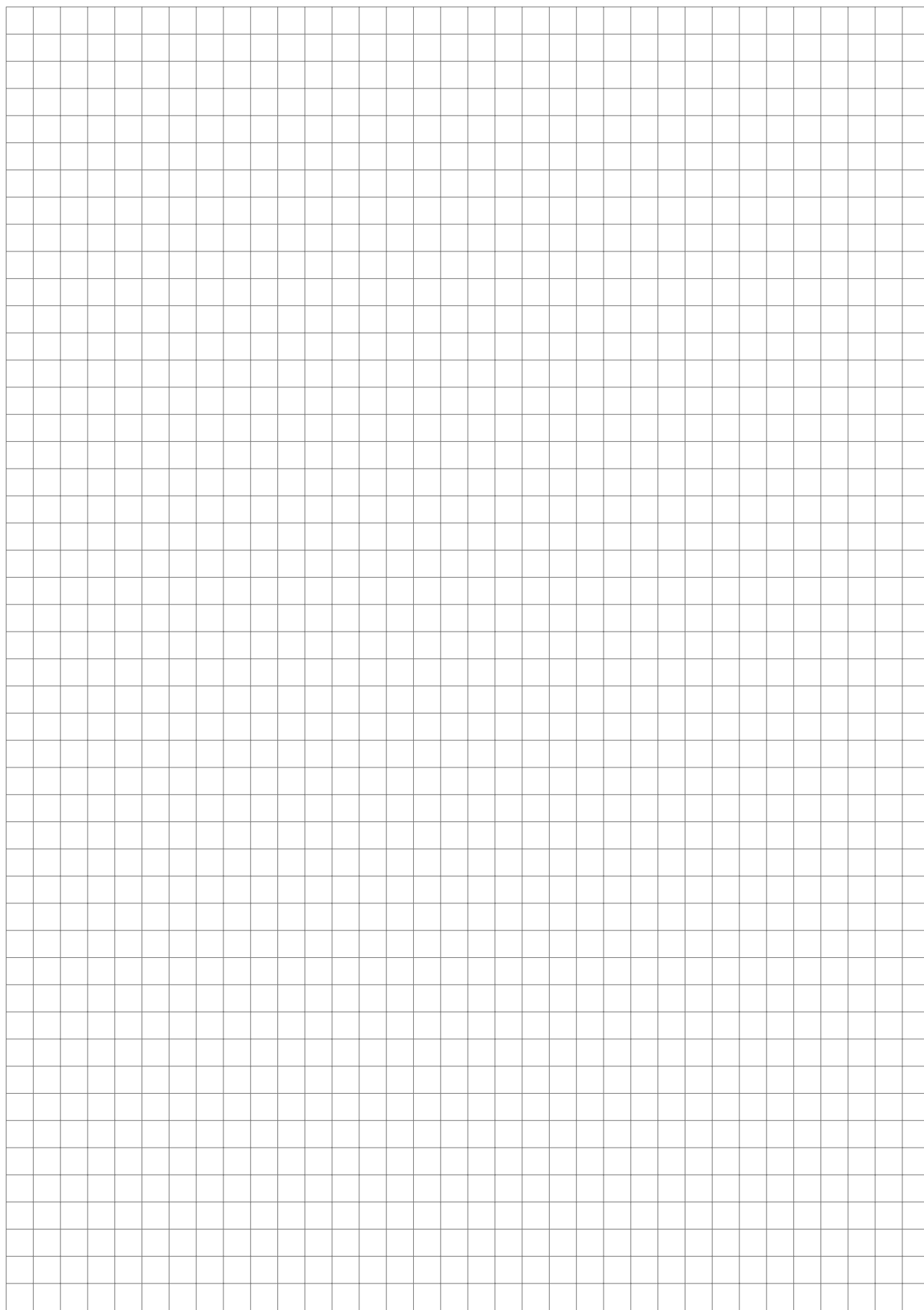
donde $Y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix}$ es un vector columna que representa n observaciones independientes, $Z = \begin{pmatrix} 1 & z_1 \\ \dots & \dots \\ 1 & z_n \end{pmatrix}$ es

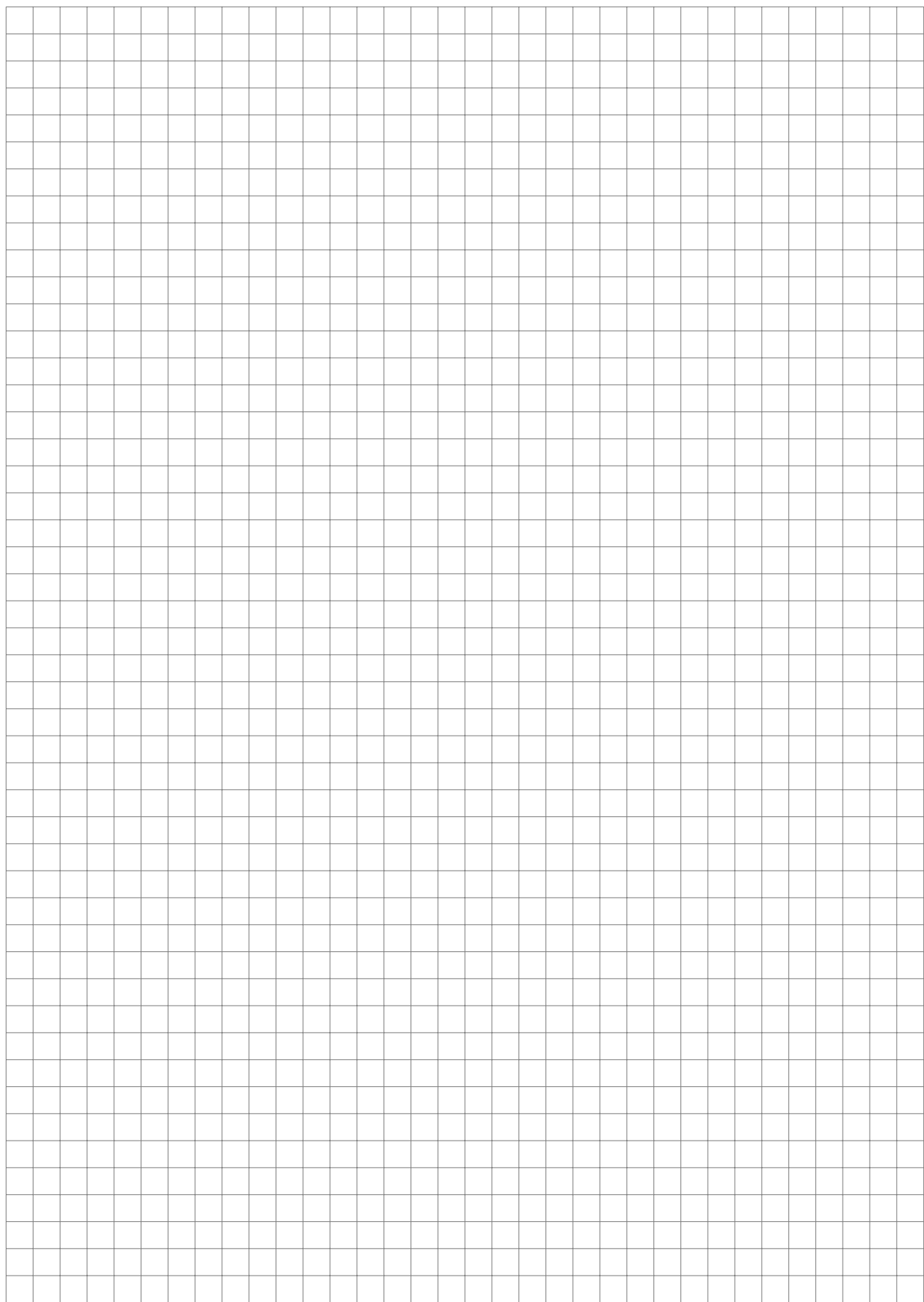
una matriz de n filas y $k = 2$ columnas, $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ es un vector de 2 parámetros, y $u = \begin{pmatrix} u_1 \\ \dots \\ u_n \end{pmatrix}$ es un vector que incluye los valores de una variable aleatoria tal que $E[u] = 0$.

1. (6 pt) Plantee el mejor estimador lineal insesgado de θ , que llamaremos $\hat{\theta}$
2. (6 pt) Derive conceptualmente $\hat{\theta}$ a partir de la suma de de los residuos cuadrados del modelo muestral (\hat{u}). Mencione el o los supuestos que sean necesarios, y exactamente en dónde le o les necesita.
3. (5 pt) Demuestre que este estimador es insesgado. Mencione el o los supuestos que sean necesarios, y exactamente en dónde le o les necesita.
4. (4 pt) Presente (no es necesario que la demuestre) la expresión para la varianza de dicho estimador, $Var(\hat{\theta}|Z)$, o al menos el valor de la primera fila $Var(\hat{\theta}_1|Z)$. ¿Qué supuesto es necesario para obtener una varianza constante de dicho estimador para todos los valores de Z ? ¿por qué?
5. (2 pt) Muestre que $\hat{\theta}$ es un estimador lineal de θ
6. (2 pt) ¿Qué otro método se puede usar para derivar $\hat{\theta}$? Plantee el sistema de ecuaciones correspondiente sin resolverlo









3. Parte III: Implementación e interpretación (15 puntos)

A continuación se presenta el código de una rutina escrita en el lenguaje R que tiene como objetivo estimar dos modelos de regresión lineal y compararlos, tal como se presenta en la Tabla 1. Los datos, que se encuentran en el archivo en formato CSV, son los resultados de un experimento en el que se le distribuyó vitaminas a un grupo de estudiantes de econometría, y dulces de menta al otro grupo. Cada fila representa un estudiante, y se incluyen las siguientes variables: 1) *trat*: es 1 si es estudiante pertenece al salón que recibió vitaminas, y 0 si es al salón que recibió mentas; 2) *hejer*: es el número de horas semanales de ejercicio reportado por cada estudiante antes del experimento; y 3) *vitalidad*: índice de vitalidad del estudiante medido en «Nairos» como unidad de medida.

Complete en los espacios indicados, ó si lo prefiere, especifique en otro lenguaje de programación la forma en la que llevaría a cabo el mismo ejercicio partiendo de la misma base de datos.

```
rm(list = ls())
library(stargazer)
setwd("C:/Users/user/data")
df = read.table("misDatos.txt", header = TRUE, sep = ";")

#[1pt] ¿Qué hacen las anteriores línea de código? _____
# _____
# _____

_____ #[1pt] Escriba el código para obtener estadísticas
# descriptivas de las tres variables

_____ #[1pt] Escriba el código para producir un histograma
# de la variable "hejer"

reg1=lm(_____, data=df) # [1pt] Complete el código para estimar
# vitalidad=b0+b1*trat+u

reg2=lm(_____, data=df) # [1pt] Complete el código para estimar
# vitalidad=a0+a1*trat+a2*hejer+u

aic1=AIC(reg1)
aic2=AIC(reg2)
stargazer(reg1, reg2, type="text", add.lines = list(c("Akaike AIC", aic1, aic2)))
```

Los resultados del ejercicio anterior se presentan a continuación. Teniéndole como referencia, responda las preguntas que se presentan después de los resultados

trat	hejer	vitalidad
Min. :0.0000	Min. :0.1516	Min. : -3.8172
1st Qu.:0.0000	1st Qu.:0.5204	1st Qu.: 0.6139
Median :1.0000	Median :0.8118	Median : 2.2309
Mean :0.5048	Mean :1.2064	Mean : 2.1793
3rd Qu.:1.0000	3rd Qu.:1.4934	3rd Qu.: 3.7093
Max. :1.0000	Max. :6.7304	Max. : 7.3520

Dependent variable:

	vitalidad	
	(1)	(2)
trat	2.858 (0.321)	2.927 (0.271)
hejer		0.792 (0.121)
Constant	0.737 (0.228)	-0.253 (0.245)
Akaike AIC	406.498304869967	371.86899572818
Observations	105	105
R2	0.435	0.601
Adjusted R2	0.429	0.593
Residual Std. Error	1.645 (df = 103)	1.389 (df = 102)
F Statistic	79.195 (df = 1; 103)	76.872 (df = 2; 102)

Note: standard errors in parenthesis

- (5pt) Interprete los coeficiente para *trat* de la columna 1, y de la columna 2. Tenga en cuenta que la variable es binaria, y el nivel de significancia de los estimadores en cada caso.
- (3pt) ¿Cuál de los dos modelos es preferido según el R^2 y el criterio *AIC* ?, ¿es importante la elección de modelo para determinar el efecto causal de las vitaminas?
- (2pt) Según la columna 2, ¿cuál es la relación entre las horas de ejercicio y la vitalidad de una persona?, ¿cree usted que el coeficiente asociado a *hejer* puede tener una interpretación causal? Explique su respuesta.

