

# Práctica Dirigida 4

Mg. Sc. J. Eduardo Gamboa U.

Los archivos que se utilizarán en esta práctica dirigida corresponden a la [Encuesta Nacional de Hogares \(ENAH\)](#), llevada a cabo por el INEI de manera trimestral desde hace más de dos décadas. Esta encuesta está dividida por módulos, cada uno de los cuales aborda una temática en particular.

Antes de iniciar con la lectura de datos, crear un proyecto y almacenar ahí los archivos de datos.

## Pregunta 1

El archivo **PD4 - datos1.txt** contiene datos acerca del módulo 85 (Gobernabilidad, Democracia y Transparencia) correspondientes al primer trimestre del año 2025.

Leer el archivo utilizando las funciones `read.table` y `read_delim`, y almacenarlo en data frames de nombres `datos1_a` y `datos1_b`, respectivamente. Luego, visualizar las 5 primeras filas de cada data frame.

```
datos1_a = read.table(file = 'PD4 - datos1.txt', header = TRUE)
str(datos1_a)
```

```
'data.frame':  7849 obs. of  9 variables:
 $ AÑO      : int  2025 2025 2025 2025 2025 2025 2025 2025 2025 ...
 $ MES      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ CONGLOME: int  15009 15009 15009 15009 15009 15009 15009 15009 15009 15024 15024 ...
 $ VIVIENDA: int  12 25 49 61 74 86 121 130 26 39 ...
 $ HOGAR    : int  11 11 11 11 11 11 11 11 11 11 11 ...
 $ CODPERSONO: int  1 1 1 1 2 2 1 1 1 1 ...
 $ CODINFOR: int  1 1 1 1 2 2 1 1 1 1 ...
 $ UBIGEO   : int  10101 10101 10101 10101 10101 10101 10101 10101 10101 10101 10101 ...
 $ DOMINIO  : int  4 4 4 4 4 4 4 4 4 4 ...
```

```
head(datos1_a, 5)
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4

```
library(readr)
datos1_b = read_delim(file = 'PD4 - datos1.txt', )
```

Rows: 7849 Columns: 9

-- Column specification -----

Delimiter: "\t"

dbl (9): AÑO, MES, CONGLOME, VIVIENDA, HOGAR, CODPERSONO, CODINFOR, UBIGEO, DO...

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
str(datos1_b)
```

```
spc_tbl_ [7,849 x 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ AÑO      : num [1:7849] 2025 2025 2025 2025 2025 ...
 $ MES      : num [1:7849] 1 1 1 1 1 1 1 1 1 1 ...
 $ CONGLOME: num [1:7849] 15009 15009 15009 15009 15009 ...
 $ VIVIENDA: num [1:7849] 12 25 49 61 74 86 121 130 26 39 ...
 $ HOGAR    : num [1:7849] 11 11 11 11 11 11 11 11 11 11 ...
 $ CODPERSONO: num [1:7849] 1 1 1 1 2 2 1 1 1 1 ...
 $ CODINFOR: num [1:7849] 1 1 1 1 2 2 1 1 1 1 ...
 $ UBIGEO   : num [1:7849] 10101 10101 10101 10101 10101 ...
 $ DOMINIO  : num [1:7849] 4 4 4 4 4 4 4 4 4 4 ...
 - attr(*, "spec")=
 .. cols(
 ..   AÑO = col_double(),
 ..   MES = col_double(),
 ..   CONGLOME = col_double(),
 ..   VIVIENDA = col_double(),
 ..   HOGAR = col_double(),
 ..   CODPERSONO = col_double(),
```

```

..   CODINFOR = col_double(),
..   UBIGEO = col_double(),
..   DOMINIO = col_double()
.. )
- attr(*, "problems")=<externalptr>

```

```
head(datos1_b,5)
```

```

# A tibble: 5 x 9
  AÑO   MES CONGLOME VIVIENDA HOGAR CODPERSO CODINFOR UBIGEO DOMINIO
<dbl> <dbl>   <dbl>   <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1  2025     1   15009     12    11         1         1   10101         4
2  2025     1   15009     25    11         1         1   10101         4
3  2025     1   15009     49    11         1         1   10101         4
4  2025     1   15009     61    11         1         1   10101         4
5  2025     1   15009     74    11         2         2   10101         4

```

## Pregunta 2

El archivo **PD4 - datos2.txt** contiene datos acerca del módulo 85 (Gobernabilidad, Democracia y Transparencia) correspondientes al primer trimestre del año 2025, pero con información añadida al inicio del archivo.

Leer las 10 primeras filas de la tabla de datos, utilizando las funciones `read.table` y `read_delim`, y almacenarlo en data frames de nombres `datos2_a` y `datos2_b`, respectivamente.

```
datos2_a = read.table(file = 'PD4 - datos2.txt', skip = 5, nrow = 10, header = TRUE)
datos2_a
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4
6	2025	1	15009	86	11	2	2	10101	4
7	2025	1	15009	121	11	1	1	10101	4
8	2025	1	15009	130	11	1	1	10101	4
9	2025	1	15024	26	11	1	1	10101	4
10	2025	1	15024	39	11	1	1	10101	4

```
datos2_b = read_delim(file = "PD4 - datos2.txt", skip = 5, n_max = 10)
datos2_b
```

# A tibble: 10 x 9

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4
6	2025	1	15009	86	11	2	2	10101	4
7	2025	1	15009	121	11	1	1	10101	4
8	2025	1	15009	130	11	1	1	10101	4
9	2025	1	15024	26	11	1	1	10101	4
10	2025	1	15024	39	11	1	1	10101	4

### Pregunta 3

El archivo **PD4 - datos3.txt** contiene datos acerca del módulo 85 (Gobernabilidad, Democracia y Transparencia) correspondientes al primer trimestre del año 2025, pero con un distinto separador de columnas.

Leer el archivo utilizando las funciones `read.table`, `read_delim`, y almacenarlo en data frames de nombres `datos3_a` y `datos3_b`, respectivamente. Luego, visualizar las 5 primeras filas de cada data frame.

```
datos3_a = read.table(file = 'PD4 - datos3.txt', header = TRUE, sep = "-",
                      encoding = "latin1")
head(datos3_a,5)
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4

```
datos3_b = read_delim(file = "PD4 - datos3.txt", delim = "-",
                      locale = locale(encoding = "Latin1"))
head(datos3_b,5)
```

# A tibble: 5 x 9

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4

Si no sabes qué encoding utilizar, puedes consultar con esta función:

```
guess_encoding("PD4 - datos3.txt")
```

# A tibble: 2 x 2

	encoding	confidence
	<chr>	<dbl>
1	ISO-8859-1	0.45
2	ISO-8859-2	0.22

## Pregunta 4

El archivo **PD4 - datos4.txt** contiene datos acerca del módulo 85 (Gobernabilidad, Democracia y Transparencia) correspondientes al primer trimestre del año 2025 en la Provincia Constitucional del Callao. Leer el archivo utilizando las funciones `read.table`, `read_delim`, y `read_tsv` y almacenarlo en data frames de nombres `datos4_a`, `datos4_b` y `datos4_c`, respectivamente. Luego, visualizar las 4 primeras filas de cada data frame.

```
# datos4_a = read.table(file = 'PD4 - datos4.txt', header = TRUE)
# ¿Por qué da error esta primera propuesta? ↑
# Corrigiendo:
datos4_a = read.table(file = 'PD4 - datos4.txt', header = TRUE, sep = "\t")
head(datos4_a,4)
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DISTRITO	P1.01
1	2025	1	16166	23	11	1	0	70104	LA PERLA	9
2	2025	1	16166	39	11	1	0	70104	LA PERLA	9
3	2025	1	16166	85	11	2	0	70104	LA PERLA	9
4	2025	1	16166	127	11	1	0	70104	LA PERLA	9

```
datos4_b = read_delim(file = "PD4 - datos4.txt")
head(datos4_b,4)
```

```
# A tibble: 4 x 10
  AÑO    MES CONGLOME VIVIENDA HOGAR CODPERSONO CODINFOR UBIGEO DISTRITO `P1$01`
<dbl> <dbl>   <dbl>   <dbl> <dbl>   <dbl>   <dbl> <dbl> <chr>      <dbl>
1  2025     1    16166     23    11       1       0  70104 LA PERLA      9
2  2025     1    16166     39    11       1       0  70104 LA PERLA      9
3  2025     1    16166     85    11       2       0  70104 LA PERLA      9
4  2025     1    16166    127    11       1       0  70104 LA PERLA      9
```

```
datos4_c = read_tsv(file = "PD4 - datos4.txt")
head(datos4_c,4)
```

```
# A tibble: 4 x 10
  AÑO    MES CONGLOME VIVIENDA HOGAR CODPERSONO CODINFOR UBIGEO DISTRITO `P1$01`
<dbl> <dbl>   <dbl>   <dbl> <dbl>   <dbl>   <dbl> <dbl> <chr>      <dbl>
1  2025     1    16166     23    11       1       0  70104 LA PERLA      9
2  2025     1    16166     39    11       1       0  70104 LA PERLA      9
3  2025     1    16166     85    11       2       0  70104 LA PERLA      9
4  2025     1    16166    127    11       1       0  70104 LA PERLA      9
```

## Pregunta 5

El archivo **PD4 - datos5.csv** contiene datos acerca del módulo 85 (Gobernabilidad, Democracia y Transparencia) correspondientes al primer trimestre del año 2025.

Leer el archivo utilizando las funciones `read.csv`, `read_csv` y `read_delim`, y almacenarlo en data frames de nombres `datos5_a`, `datos5_b` y `datos5_c`, respectivamente. Luego, visualizar las 4 primeras filas de cada data frame.

```
datos5_a = read.csv('PD4 - datos5.csv', encoding = "latin1")
head(datos5_a, 4)
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4

```
datos5_b = read_csv('PD4 - datos5.csv', locale = locale(encoding = "Latin1"))
head(datos5_b, 4)
```

# A tibble: 4 x 9

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4

```
datos5_c = read_delim('PD4 - datos5.csv', delim = ",",
                      locale = locale(encoding = "Latin1"))
head(datos5_c, 4)
```

# A tibble: 4 x 9

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4

## Pregunta 6

El archivo **PD4 - datos6.csv** contiene datos acerca del módulo 85 (Gobernabilidad, Democracia y Transparencia) correspondientes al primer trimestre del año 2025.

Leer el archivo utilizando las funciones `read.csv2`, `read.csv`, `read.table`, `read_csv2` y `read_delim`, y almacenarlo en data frames de nombres `datos6_a`, `datos6_b`, `datos6_c`, `datos6_d` y `datos6_e` respectivamente. Luego, visualizar las 5 primeras filas de cada data frame.

```
datos6_a = read.csv2('PD4 - datos6.csv', encoding = "latin1")
head(datos6_a)
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4
6	2025	1	15009	86	11	2	2	10101	4

```
datos6_b = read.csv('PD4 - datos6.csv', sep = ";", encoding = "latin1")
head(datos6_b)
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4
6	2025	1	15009	86	11	2	2	10101	4

```
datos6_c = read.table('PD4 - datos6.csv', sep = ";", encoding = "latin1")
head(datos6_c)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
2	2025	1	15009	12	11	1	1	10101	4
3	2025	1	15009	25	11	1	1	10101	4
4	2025	1	15009	49	11	1	1	10101	4
5	2025	1	15009	61	11	1	1	10101	4
6	2025	1	15009	74	11	2	2	10101	4



```
datos6_d = read_csv2('PD4 - datos6.csv', locale = locale(encoding = "Latin1"))
head(datos6_d)
```

```
# A tibble: 6 x 9
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4
6	2025	1	15009	86	11	2	2	10101	4

```
datos6_e = read_delim('PD4 - datos6.csv', delim = ";",
                      locale = locale(encoding = "Latin1"))
head(datos6_e)
```

```
# A tibble: 6 x 9
```

	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	CODINFOR	UBIGEO	DOMINIO
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2025	1	15009	12	11	1	1	10101	4
2	2025	1	15009	25	11	1	1	10101	4
3	2025	1	15009	49	11	1	1	10101	4
4	2025	1	15009	61	11	1	1	10101	4
5	2025	1	15009	74	11	2	2	10101	4
6	2025	1	15009	86	11	2	2	10101	4