

Economía del Conocimiento, TICs e Innovación

Guía de replicación

Provincia de Misiones

2018-2022



AUTORIDADES

GOBERNADOR

HUGO PASSALACQUA

VICEGOBERNADOR

LUCAS ROMERO SPINELLI

DIRECTORA EJECUTIVA INSTITUTO PROVINCIAL DE ESTADÍSTICA Y CENSOS

SILVANA DEA LABAT

Contenido

1. INTRODUCCIÓN	4
2. EL ARCHIVO DE MICRODATOS.....	5
2.1. Estructura general	5
2.2. Diccionario de variables	5
2.3. Tipos de empresa según el diseño muestral	7
3. LA FÓRMULA FUNDAMENTAL DE ESTIMACIÓN	8
4. MAPA DE VARIABLES POR CUADRO	9
5. REPLICA: CUADRO POR CUADRO.....	10
6. VERIFICACIONES DE CONSISTENCIA INTERNA	13
7. SCRIPT COMPLETO DE REPLICA.....	14
8. PREGUNTAS FRECUENTES	16

1. INTRODUCCIÓN

Este documento explica, paso a paso, cómo reproducir cada uno de los 12 (doce) cuadros estadísticos publicados en el informe oficial Economía del Conocimiento, TICs e Innovación de la Provincia de Misiones utilizando exclusivamente el archivo de microdatos MICRODATOS_ECTI_MISIONES_REPOSITORIO.csv.

El propósito central reside en garantizar la transparencia y la auditabilidad de las estadísticas provinciales, permitiendo que cualquier usuario pueda verificar de manera independiente las cifras publicadas.

La guía adopta un enfoque pedagógico: comienza con los conceptos fundamentales, presenta la fórmula general de estimación, y luego recorre cuadro por cuadro las instrucciones específicas con ejemplos de código reproducible en Python y los resultados esperados. No se requieren conocimientos previos de muestreo estadístico para seguir las instrucciones, aunque se ofrece contexto metodológico para quienes deseen profundizar.

2. EL ARCHIVO DE MICRODATOS

2.1. Estructura general

El archivo contiene 95 registros (filas) y 46 variables (columnas). Cada registro representa una empresa encuestada. Las empresas están clasificadas según las secciones de la Clasificación Industrial Internacional Uniforme (CIIU) Revisión 4, identificadas con letras de la A a la T.

Las variables de innovación y TICs son binarias: 1 indica que la empresa realizó la actividad o posee el recurso; 0 indica que no; y un valor faltante (NaN) indica que la empresa no respondió esa pregunta.

2.2. Diccionario de variables

Cuadro 1. Variables de identificación y ponderación. Provincia de Misiones. Periodo año 2018-2022

Variable	Tipo de variable	Descripción
codigo_empresa	Texto	Código anónimo único (ECTI-NNNN)
sector	Texto	Sección CIIU Rev. 4 (A, C, G, etc.)
tamaño	Texto	Micro
		Pequeña
		Media T1
		Media T2
		Grande
tipo_muestra	Texto	PROB
		IF
		REEMP
		ATM_EXTRA
		SIN_MARCO
pondera	Numérica	Ponderador calibrado original
pondera_trim	Numérica	Ponderador con trimming (máximo 500)
es_innovadora	Binaria	1 = realizó al menos 1 actividad innovativa
tiene_vinculacion	Binaria	1 = se vinculó con otro agente para innovar

Fuente. IPEC, elaborado en base a la Encuesta PBG, Economía Circular, Cambio Climático y Economía del Conocimiento en la Era Digital (2023-2024) y registros administrativos de la Agencia Tributaria de Misiones (ATM).

Cuadro 2. Variables de innovación. Provincia de Misiones. Periodo año 2018-2022

Variable	Descripción
innov_id_interno	Investigación y desarrollo interno
innov_id_externo	Investigación y desarrollo externo
innov_diseno_interno	Diseño industrial / ingeniería (interno)
innov_diseno_externo	Diseño industrial / ingeniería (externo)
innov_marketing_int	Marketing para innovación (interno)
innov_marketing_ext	Marketing para innovación (externo)
innov_maquinaria	Adquisición de maquinaria y equipos
innov_hardware_sw	Adquisición de hardware y software
innov_edificios	Edificios y terrenos para innovación
innov_conoc_externos	Adquisición de conocimientos externos
innov_capacitacion	Capacitación de trabajadores para innovación

Fuente. IPEC, elaborado en base a la Encuesta PBG, Economía Circular, Cambio Climático y Economía del Conocimiento en la Era Digital (2023-2024) y registros administrativos de la Agencia Tributaria de Misiones (ATM).

Cuadro 3. Variables TICs. Provincia de Misiones. Periodo año 2018-2022

Variable	Descripción
tic_internet	Usa servicio de internet
tic_pc	Posee ordenadores / PC
tic_lan	Posee red de área local (LAN)
tic_telefonia_movil	Telefonía móvil empresarial
tic_gestion_admin	Sistema de gestión administrativa
tic_firma_digital	Usa firma digital
tic_cloud	Usa cloud computing
tic_pagina_web	Posee página web propia
tic_pedidos_web	Recibe pedidos/reservas por web
tic_especialista	Tiene especialistas TICs en planta
tic_whatsapp	Usa WhatsApp comercial
tic_instagram	Usa Instagram
tic_facebook	Usa Facebook

Fuente. IPEC, elaborado en base a la Encuesta PBG, Economía Circular, Cambio Climático y Economía del Conocimiento en la Era Digital (2023-2024) y registros administrativos de la Agencia Tributaria de Misiones (ATM).

2.3. Tipos de empresa según el diseño muestral

Las empresas se clasifican en cinco categorías según su origen en el diseño muestral:

1. PROB (48 empresas) provienen de la muestra probabilística y tienen ponderadores calibrados que representan a múltiples empresas del universo.
2. IF (10 empresas) son de inclusión forzosa, con ponderador 1,0 y se representan a sí mismas.
3. Reemplazos (25 empresas) heredan el ponderador de la empresa original
4. ATM_EXTRA (15 empresas) fueron identificadas en el padrón de la Agencia Tributaria de Misiones y tienen ponderador 1,0
5. SIN_MARCO (2 empresas) no pudieron identificarse en ningún marco

3. LA FÓRMULA FUNDAMENTAL DE ESTIMACIÓN

3.1. La regla de oro

Toda la réplica del informe descansa sobre una única operación aritmética. Para obtener cualquier tasa publicada en el informe, se aplica:

$$\hat{t} = \frac{\sum (y_i * pondera_trim_i)}{\sum pondera_trim_i} * 100$$

donde la sumatoria recorre todos los registros i que tienen respuesta válida (no NaN) en la variable y . El resultado se expresa como porcentaje del universo estimado de empresas.

3.2. ¿Por qué se multiplica por pondera_trim?

La encuesta ECTI no relevó a todas las empresas de Misiones, sino a una muestra de 95 empresas. El ponderador de cada empresa indica cuántas empresas del universo real ella representa. Al multiplicar el indicador binario de una empresa por su ponderador, se expande la información muestral hacia el universo completo. Las empresas de inclusión forzosa tienen ponderador 1,0 porque se representan solo a sí mismas. Las probabilísticas tienen ponderadores mayores, porque cada una representa a muchas empresas similares no encuestadas.

3.3. ¿Por qué pondera_trim y no pondera?

El ponderador original *pondera* puede alcanzar valores de hasta 3.761 lo que genera inestabilidad en las estimaciones. El ponderador recortado *pondera_trim* impone un máximo de 500, siguiendo las recomendaciones de la literatura estadística sobre estabilización de pesos extremos. Las cifras del informe utilizan *pondera_trim* como ponderador por defecto.

3.4. La fórmula en Python

En su forma más concreta, la operación se escribe así:

```
import pandas as pd

df = pd.read_csv('MICRODATOS_ECTI_MISIONES_REPOSITORIO.csv')

def tasa(df, variable):
    m = df[variable].notna()
    return (df.loc[m, variable] * df.loc[m, 'pondera_trim']).sum() \
        / df.loc[m, 'pondera_trim'].sum() * 100
```

Esta función permite replicar la totalidad de los cuadros del informe. Lo que varía de un cuadro a otro es la variable que se usa y cómo se agrupan los resultados.

4. MAPA DE VARIABLES POR CUADRO

Cuadro 4. Correspondencia entre cuadros del informe y variables de los microdatos. Provincia de Misiones.
Periodo año 2018-2022

Cuadro	Contenido	Variables	Agrupación
1	Actividades de innovación	innov_* (11 vars)	Global
2	Innovación por tamaño	es_innovadora	Por tamaño
3	Vinculación I+D	vinc_* (7 vars)	Global
4	Socios para la innovación	socio_* (9 vars)	Global
5	Brecha innovadora	es_innovadora (filtro)	Subgrupos
6	Equipamiento TICs	tic_internet, tic_pc, tic_lan	Global
7	Herramientas avanzadas	tic_gestion_admin, tic_firma_digital, ...	Global
8	Procesos automatizados	(fuente directa)	Global
9	Canales digitales	tic_whatsapp, tic_instagram, ...	Global
10	Medios de pago	(fuente directa)	Global
11	TICs x innovación	tic_* x es_innovadora	Cruzado
12	Índice TICs por tamaño	indice_tic	Por tamaño

Fuente. IPEC, elaborado en base a la Encuesta PBG, Economía Circular, Cambio Climático y Economía del Conocimiento en la Era Digital (2023-2024) y registros administrativos de la Agencia Tributaria de Misiones (ATM).

5. REPLICA: CUADRO POR CUADRO

5.1. Cuadro 1: actividades de innovación

Este cuadro presenta el porcentaje de empresas que realizaron cada tipo de actividad innovativa. Para replicarlo, se aplica la fórmula fundamental a cada una de las 11 variables innovativas

```
# Cuadro 1: Actividades de innovación

innov_vars = [c for c in df.columns if c.startswith('innov_')]

for var in innov_vars:

    print(f'{var}: {tasa(df, var):.1f}%')
```

Valores esperados: innov_maquinaria = 11,9%; innov_hardware_sw = 1,9%; innov_capacitacion = 1,3%; innov_id_interno = 0,4%.

5.2. Cuadro 2: innovación por tamaño

Se agrupa las actividades innovativas por la variable tamaño y se calcula la tasa de innovación dentro de cada grupo.

```
# Cuadro 2: Innovación por tamaño

for tam in ['Micro','Pequeña','Media T1','Media T2','Grande']:

    sub = df[df['tamaño'] == tam]

    t = tasa(sub, 'es_innovadora')

    print(f'{tam}: {t:.1f}%')
```

Valores esperados: Micro = 6,8%; Pequeña = 8,3%; Media T1 = 39,9%; Media T2 = 52,8%; Grande = 78,3%.

5.3. Cuadro 3: vinculación para innovación

Análogo al cuadro 1, aplicando la fórmula a las 7 (siete) variables de vinculación para la innovación

```
# Cuadro 3: Vinculación

vinc_vars = [c for c in df.columns if c.startswith('vinc_')]

for var in vinc_vars:

    print(f'{var}: {tasa(df, var):.1f}%')
```

Valores esperados: vinc_capacitacion_rrhh = 12,7%; vinc_cambios_org = 12,4%; vinc_id = 10,3%.

5.4. Cuadro 4: socios para innovación

Análogo al cuadro anterior, aplicando la fórmula a las 9 (nueve) variables de socio para la innovación

```
# Cuadro 4: Socios
socio_vars = [c for c in df.columns if c.startswith('socio_')]

for var in socio_vars:
    print(f'{var}: {tasa(df, var):.1f}%')
```

Valor esperado: socio_clientes = 6,9%; socio_universidades = 0,0%.

5.5. Cuadro 5: brecha empresas innovadoras y no innovadoras

Se separa el dataset en dos subgrupos y se comparan los indicadores. Los ratios de facturación y empleo no son replicables desde los microdatos porque la facturación (fuente ATM) no se incluye por confidencialidad.

5.6. Cuadros 6 y 7: infraestructura y adopción TICs

```
# Cuadros 6-7: TICs
tic_vars = [c for c in df.columns if c.startswith('tic_')]

for var in tic_vars:
    print(f'{var}: {tasa(df, var):.1f}%')
```

Valores esperados: tic_internet = 97,0%; tic_pc = 93,0%; tic_gestion_admin = 75,0%.

5.7. Cuadro 11: TICs según condición innovadora

Se divide el dataset por la condición, si la empresa es o no innovadora y se calcula la tasa de cada TICs en cada subgrupo.

```
# Cuadro 11: TICs × innovación
inn = df[df['es_innovadora'] == 1]
no_inn = df[df['es_innovadora'] == 0]

tic_vars = ['tic_especialista', 'tic_firma_digital', 'tic_pagina_web',
            'tic_gestion_admin', 'tic_whatsapp']

for var in tic_vars:
    ti = tasa(inn, var)
    tn = tasa(no_inn, var)
    print(f'{var}: inn={ti:.1f}% | no_inn={tn:.1f}% | Δ={ti-tn:+.1f}%')
```

Valor esperado: tic_especialista: inn = 72,2% | no_inn = 8,6% | Δ = 63,7

5.8. Cuadro 12: Índice TICs por tamaño

Cuadro 12: Índice TICs por tamaño

```
for tam in ['Micro','Pequeña','Media T1','Media T2','Grande']:
```

```
    sub = df[df['tamaño'] == tam]
```

```
    idx = sub['indice_tic'].dropna()
```

```
    print(f'{tam}: media={idx.mean():.1f}')
```

Valores esperados: Micro = 2,9; Pequeña = 3,4; Media T1 = 4,5; Media T2 = 5,9; Grande = 5,8

6. VERIFICACIONES DE CONSISTENCIA INTERNA

Estas pruebas no requieren datos externos: se basan íntegramente en la estructura lógica del archivo de microdatos.

6.1. Identidad es_innovadora = max(innov_*)

Para cada empresa, es_innovadora debe valer 1 si y solo si al menos una variable innov_* vale 1:

```
innov_cols = [c for c in df.columns if c.startswith('innov_')]
check = (df['es_innovadora'] == (df[innov_cols].max(axis=1) > 0).astype(int))
print(f'Consistencia: {check.mean()*100:.0f}%') # → 100%
```

6.2. Identidad indice_tic = suma de componentes

```
tic_idx = ['tic_internet', 'tic_pc', 'tic_lan', 'tic_gestion_admin',
           'tic_firma_digital', 'tic_cloud', 'tic_pagina_web', 'tic_whatsapp']
check = (df['indice_tic'] - df[tic_idx].sum(axis=1)).abs().max()
print(f'Error máximo: {check:.6f}') # → 0.000000
```

6.3. Universo estimado

```
universo = df['pondera_trim'].sum()
print(f'Universo: {universo:.0f}') # → ~4.883
```

6.4. Tasa de innovación

```
tasa_inn = tasa(df, 'es_innovadora')
print(f'Tasa innovación: {tasa_inn:.1f}%') # → 13.5%
```

7. SCRIPT COMPLETO DE REPLICA

A continuación, se presenta un script Python autocontenido que replica todos los cuadros del informe y ejecuta las verificaciones de consistencia. Puede copiarse y ejecutarse directamente sobre el archivo de microdatos.

```
import pandas as pd

df = pd.read_csv('MICRODATOS_ECTI_MISIONES_REPOSITORIO.csv')

def tasa(df, variable):
    m = df[variable].notna()
    if m.sum() == 0: return 0.0
    return (df.loc[m, variable] * df.loc[m, 'pondera_trim']).sum() \
        / df.loc[m, 'pondera_trim'].sum() * 100

# === CUADRO 1: Actividades de innovación ===
print('=== CUADRO 1 ===')
for c in [c for c in df.columns if c.startswith('innov_')]:
    print(f' {c}: {tasa(df, c):.1f}%')

# === CUADRO 2: Innovación por tamaño ===
print('\n=== CUADRO 2 ===')
for t in ['Micro', 'Pequeña', 'Media T1', 'Media T2', 'Grande']:
    print(f' {t}: {tasa(df[df["tamaño"]==t, "es_innovadora"]):.1f}%')

# === CUADRO 3: Vinculación ===
print('\n=== CUADRO 3 ===')
for c in [c for c in df.columns if c.startswith('vinc_')]:
    print(f' {c}: {tasa(df, c):.1f}%')

# === CUADRO 4: Socios ===
print('\n=== CUADRO 4 ===')
for c in [c for c in df.columns if c.startswith('socio_')]:
    print(f' {c}: {tasa(df, c):.1f}%')
```

```

# === CUADROS 6-7: TICs ===
print('\n=== CUADROS 6-7 ===')
for c in [c for c in df.columns if c.startswith('tic_')]:
    print(f' {c}: {tasa(df, c):.1f}%')

# === CUADRO 11: TICs × innovación ===
print('\n=== CUADRO 11 ===')
inn = df[df['es_innovadora']==1]
no_inn = df[df['es_innovadora']==0]
for c in [c for c in df.columns if c.startswith('tic_')
          and c != 'indice_tic']:
    ti, tn = tasa(inn, c), tasa(no_inn, c)
    print(f' {c}: inn={ti:.1f}% no={tn:.1f}% D={ti-tn:+.1f}')

# === CUADRO 12: Índice TIC por tamaño ===
print('\n=== CUADRO 12 ===')
for t in ['Micro', 'Pequeña', 'Media T1', 'Media T2', 'Grande']:
    idx = df[df['tamaño']==t]['indice_tic'].dropna()
    print(f' {t}: media={idx.mean():.1f}')

# === VERIFICACIONES ===
print('\n=== VERIFICACIONES ===')
ic = [c for c in df.columns if c.startswith('innov_')]
v1 = (df['es_innovadora']==(df[ic].max(axis=1)>0).astype(int)).all()
print(f' V1 es_innovadora consistente: {v1}')
v2 = (df['pondera_trim'] <= 500).all()
print(f' V2 trimming respetado: {v2}')
v3 = df['pondera_trim'].sum()
print(f' V3 universo: {v3:,.0f}')
v4 = tasa(df, 'es_innovadora')
print(f' V4 tasa innovación: {v4:.1f}%')
print('\n¡Replicación completa!')

```

8. PREGUNTAS FRECUENTES

¿Por qué algunos ponderadores son menores que 1?

Los ponderadores de las empresas probabilísticas fueron calibrados mediante un estimador GREG (estimador de regresión generalizada). Este proceso puede producir ponderadores menores que 1 en empresas cuya facturación declarada supera ampliamente la facturación media de su estrato en el marco ATM.

¿Por qué no se incluye la facturación ni el CUIT?

Por razones de confidencialidad estadística (Ley 17.622). La facturación proviene de registros administrativos tributarios de la ATM y su divulgación a nivel de empresa individual no está autorizada. El código ECTI-NNNN reemplaza al CUIT como identificador anónimo.

¿Qué significa que una variable TICs tenga valor NaN?

Significa que la empresa no respondió esa pregunta del cuestionario. Esto puede deberse a que la pregunta solo aplica a un tipo de persona (humana o jurídica) y la empresa es del otro tipo, o a que omitió la respuesta. Al calcular tasas con la fórmula fundamental, los registros con NaN se excluyen automáticamente tanto del numerador como del denominador.

¿Por qué la tasa de innovación cambia si uso pondera en vez de pondera_trim?

Con el ponderador original sin recortar (pondera), ocho microempresas no innovadoras con ponderadores superiores a 500 inflan artificialmente el denominador, reduciendo la tasa del 13,5% al 7,8%. El informe utiliza pondera_trim como estimación central y reporta la estimación sin trimming como referencia de sensibilidad.

¿Puedo usar estos microdatos en R?

Sí. El archivo CSV se importa directamente con `read.csv()` y la fórmula de estimación se implementa con `weighted.mean()`:

```
df <- read.csv('MICRODATOS_ECTI_MISIONES_REPOSITORIO.csv')  
# Tasa de innovación:  
weighted.mean(df$es_innovadora, df$pondera_trim, na.rm=TRUE) * 100
```

¿Puedo replicar el cuadro 5 desde estos microdatos?

Parcialmente. Los ratios de brecha requieren la facturación y el empleo, que no se incluyen en los microdatos públicos por confidencialidad. Sin embargo, puede verificarse la composición de los grupos (N innovadoras = 26; N no innovadoras = 69) filtrando por `es_innovadora`.

Economía del conocimiento, TICs e Innovación | Guía de replicación

Provincia de Misiones | 2018-2022

MISIONES
G O B I E R N O

IPEC
Instituto Provincial
de Estadística y Censos

25 de Mayo 1460 2° y 3° piso

Tel: 0376 - 4447109

Web: www.ipec.misiones.gov.ar

Email: ipec@misiones.gov.ar

CP 3300 Posadas - Misiones