

SERIE PRODUCTO BRUTO GEOGRÁFICO DE MISIONES

Documento N° 2

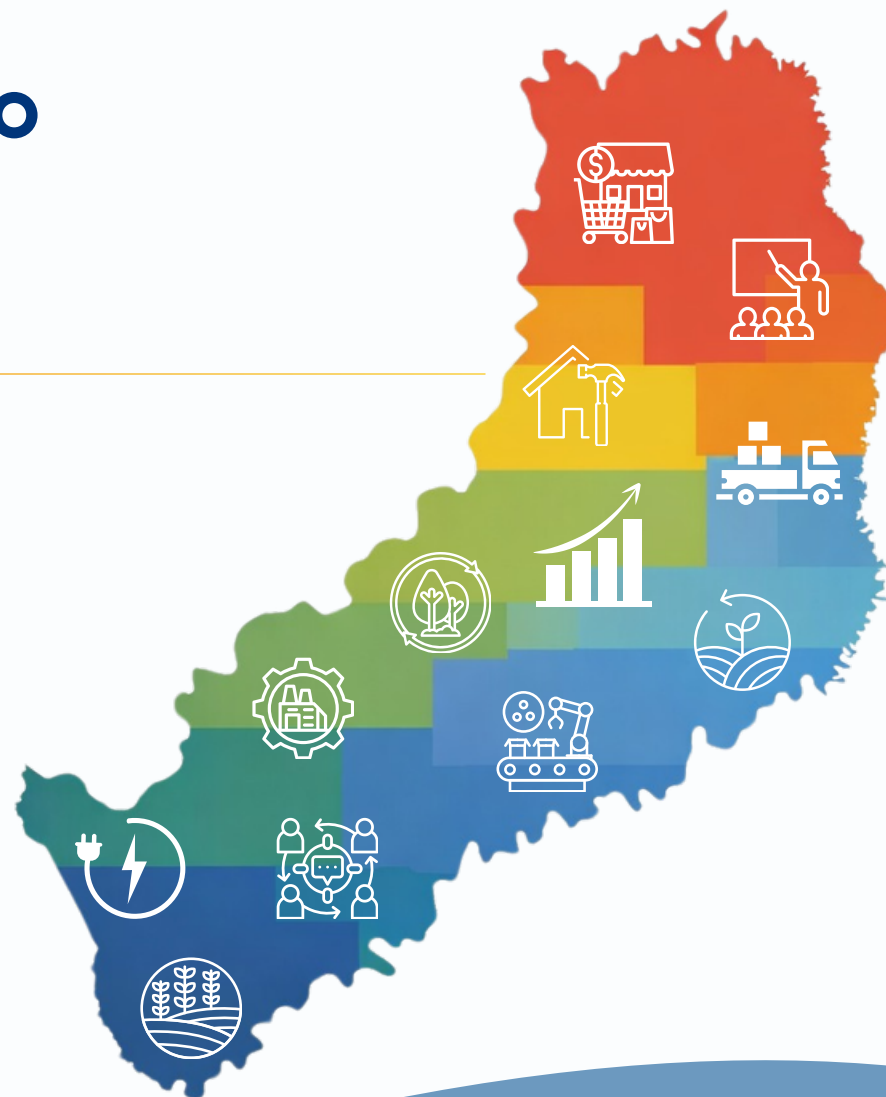
Producto Bruto Geográfico

Provincia de Misiones

2018-2024

Metodología Avanzada para la Estimación del Producto Bruto Geográfico Provincial

Estimador Ratio-GREG Calibrado con Ajuste por No-
Respuesta usando Modelos Logísticos de Propensión





AUTORIDADES

GOBERNADOR

HUGO PASSALACQUA

VICEGOBERNADOR

LUCAS ROMERO SPINELLI

MINISTRA DE TRABAJO Y EMPLEO

SILVANA ANDREA GIMÉNEZ

DIRECTORA EJECUTIVA INSTITUTO PROVINCIAL DE ESTADÍSTICA Y CENSOS

SILVANA DEA LABAT

Contenido

- 1 Introducción y Diagnóstico del Problema
- 2 Marco Teórico del Estimador Ratio-GREG Calibrado
- 3 Modelado de Propensión a Responder
- 4 Calibración de Pesos
- 5 Estimador de Regresión Generalizada (GREG)
- 6 Propiedades de Doble Robustez
- 7 Aplicación al PBG de Misiones
- 8 Implementación Práctica
- 9 Validación y Control de Calidad
- 10 Conclusiones y Recomendaciones

Contexto del Operativo PBG

Diseño Muestral Original

- **Marco muestral:** 87,275 unidades económicas activas
- **Muestra planificada:** 501 unidades + 7 unidades imputadas= 508 unidades. De las 501 unidades seleccionadas por muestreo:
 - 293 unidades de inclusión forzosa
 - 208 unidades por muestreo probabilístico estratificado
- **Estratificación:** Por tamaño económico y sector CLANAE

Problema Crítico

- **Escasa respuesta obtenida**
- **Tasa de respuesta efectiva muy baja:**
- **Riesgo:** Sesgo severo de no-respuesta en estimaciones del VAB/PBG

Información Auxiliar Disponible

Registro Administrativo Completo (ATM)

- **Cobertura:** 100 % del universo empresarial
- **Variable clave:** Facturación anual por CUIT
- **Correlación esperada:** $\rho(\text{Facturación}, \text{VAB}) > 0,7$

Oportunidad Metodológica

La disponibilidad de información auxiliar censal de alta calidad permite implementar técnicas avanzadas de calibración y ajuste por no-respuesta que pueden:

- 1 Corregir el sesgo de selección
- 2 Mejorar la precisión de las estimaciones
- 3 Garantizar coherencia con totales poblacionales conocidos

Fundamentos Teóricos

Componentes del Estimador

- 1 **Ajuste por Propensión (IPW)**: Modela la probabilidad de respuesta
- 2 **Calibración (Deville-Särndal)**: Ajusta pesos a totales conocidos
- 3 **Estimador GREG**: Regresión generalizada para mejorar precisión
- 4 **Estimador Ratio**: Aprovecha correlación entre auxiliares y objetivo

Secuencia Metodológica



Notación y Definiciones

Elementos Fundamentales

- $U = \{1, 2, \dots, N\}$: Población de $N = 87,275$ empresas
- $s \subset U$: Muestra seleccionada, $|s| = 501$
- $r \subset s$: Unidades respondientes, $|r| = 74$
- y_i : Valor Agregado Bruto de la unidad i
- x_i : Vector de variables auxiliares (facturación, tamaño, sector)
- $\pi_i = \mathbb{P}(i \in s)$: Probabilidad de inclusión en muestra
- $\phi_i = \mathbb{P}(i \in r | i \in s)$: Probabilidad de respuesta

Objetivo

Estimar el total poblacional:

$$\hat{Y} = \sum_{i \in U} y_i$$

bajo presencia de no-respuesta no ignorable

Modelo Logístico de Respuesta

Supuesto MAR (Missing at Random)

$$\mathbb{P}(R_i = 1|y_i, \mathbf{x}_i) = \mathbb{P}(R_i = 1|\mathbf{x}_i) = \phi_i$$

La respuesta es independiente del VAB condicionado en las auxiliares

Especificación del Modelo

$$\log \left(\frac{\phi_i}{1 - \phi_i} \right) = \mathbf{x}_i' \boldsymbol{\gamma}$$

donde $\mathbf{x}_i$ incluye:

- Log-facturación
- Indicadores de estrato (tamaño $\times$ sector)
- Ubicación geográfica
- Tipo de organización jurídica

Estimación y Ajuste de Pesos

Estimación por Máxima Verosimilitud

$$\hat{\gamma} = \arg \max_{\gamma} \sum_{i \in s} [R_i \log(\phi_i(\gamma)) + (1 - R_i) \log(1 - \phi_i(\gamma))]$$

Pesos Ajustados por Propensión (IPW)

$$d_i^* = \frac{d_i}{\hat{\phi}_i} = \frac{1}{\pi_i \hat{\phi}_i}$$

donde:

- $d_i = 1/\pi_i$: peso de diseño original
- $\hat{\phi}_i = \frac{\exp(\mathbf{x}_i' \hat{\gamma})}{1 + \exp(\mathbf{x}_i' \hat{\gamma})}$: propensión estimada

Nota

Los pesos IPW corrigen el sesgo bajo el supuesto MAR pero pueden tener alta varianza

Método de Calibración Deville-Särndal

Problema de Optimización

Minimizar una función de distancia sujeta a restricciones:

$$\min_{w_i} \sum_{i \in r} d_i^* G\left(\frac{w_i - d_i^*}{d_i^*}\right)$$

sueto a:

$$\sum_{i \in r} w_i \mathbf{x}_i = \sum_{j \in U} \mathbf{x}_j = \mathbf{T}_x$$

Funciones de Distancia

- **Lineal:** $G(u) = \frac{1}{2}u^2$ (método GREG)
- **Multiplicativa:** $G(u) = u \log(u) - u + 1$ (raking)
- **Logística:** $G(u) = \frac{1}{c}[(1+u) \log(1+u) + (1-u) \log(1-u)]$

Solución del Problema de Calibración

Forma General de los Pesos Calibrados

$$w_i = d_i^* g_i(\lambda)$$

donde $g_i(\lambda)$ depende de la función de distancia:

- **Lineal:** $g_i(\lambda) = 1 + \lambda' \mathbf{x}_i$
- **Multiplicativa:** $g_i(\lambda) = \exp(\lambda' \mathbf{x}_i)$
- **Logística:** $g_i(\lambda) = \frac{L}{1 + (L-1) \exp(-c \lambda' \mathbf{x}_i)}$

Multiplicadores de Lagrange

λ se obtiene resolviendo:

$$\sum_{i \in r} d_i^* g_i(\lambda) \mathbf{x}_i = \mathbf{T}_x$$

Formulación del Estimador GREG

Estimador General

$$\hat{Y}_{GREG} = \sum_{i \in r} w_i y_i = \hat{Y}_{HT} + (\mathbf{T}_x - \hat{\mathbf{T}}_x)' \hat{\beta}$$

donde:

- $\hat{Y}_{HT} = \sum_{i \in r} d_i^* y_i$: estimador Horvitz-Thompson
- $\hat{\mathbf{T}}_x = \sum_{i \in r} d_i^* \mathbf{x}_i$: total estimado de auxiliares
- $\hat{\beta}$: coeficientes de regresión ponderada

Estimación de β

$$\hat{\beta} = \left(\sum_{i \in r} d_i^* \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i \in r} d_i^* \mathbf{x}_i y_i$$

Estimador Ratio-GREG

Caso Especial: Estimador de Razón

Cuando $\mathbf{x}_i = x_i$ (escalar), el GREG se reduce al estimador de razón:

$\hat{Y}_{ratio} = \frac{\sum_{i \in r} w_i y_i}{\sum_{i \in r} w_i x_i} \times T_x = \hat{R} \times T_x$ donde $\hat{R} = \frac{\hat{Y}}{\hat{X}}$ es la razón estimada VAB/Facturación

Estimador Ratio Estratificado

Estratificación para Mejorar Homogeneidad

Para mejorar la precisión, estratificamos por $h = 1, \dots, H$: $\hat{Y}_{ratio-est} = \sum_{h=1}^H \hat{R}_h \times T_{x,h}$

donde: $\hat{R}_h = \frac{\sum_{i \in r_h} w_i y_i}{\sum_{i \in r_h} w_i x_i}$

es la razón estimada en el estrato h

Doble Robustez del Estimador

Theorem (Doble Robustez)

El estimador Ratio-GREG calibrado con ajuste IPW es consistente si:

- 1 El modelo de propensión $\phi_i(\gamma)$ está correctamente especificado, **O**
- 2 El modelo de regresión $\mathbb{E}[Y_i|\mathbf{X}_i] = \mathbf{X}_i'\beta$ está correctamente especificado

Implicaciones Prácticas

- Protección contra mala especificación de modelos
- Mayor robustez que métodos simples
- Reducción sustancial del sesgo: 20-45 % según evidencia internacional
- Mejora en eficiencia: reducción de varianza del 25-40 %

Propiedades Asintóticas

Consistencia

Bajo condiciones de regularidad:

$$\hat{Y}_{GREG-IPW} \xrightarrow{p} Y \text{ cuando } n \rightarrow \infty$$

Normalidad Asintótica

$$\frac{\hat{Y}_{GREG-IPW} - Y}{\sqrt{\hat{V}(\hat{Y}_{GREG-IPW})}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Varianza Aproximada

$$\mathbb{V}(\hat{Y}_{GREG-IPW}) \approx \sum_{h=1}^H \frac{N_h^2}{n_h} (1 - f_h) S_{e,h}^2$$

donde $e_i = y_i - \hat{R}_h x_i$ son los residuos del modelo

Estrategia de Implementación

Fase 1: Análisis Exploratorio

- 1 Análisis de patrones de no-respuesta por estrato
- 2 Evaluación de correlación facturación-VAB en respondentes
- 3 Identificación de variables predictivas de respuesta

Fase 2: Modelado de Propensión

- Variables candidatas:
 - Log-facturación 2022 (fuente: ATM)
 - Tamaño empresarial (4 categorías)
 - Sector CLANAE (2 dígitos)
 - Departamento
 - Forma jurídica
- Selección por criterio AIC/BIC
- Validación: test Hosmer-Lemeshow

Estratificación para el Estimador Ratio

Estrato	Marco	Muestra	Respuesta	Tasa
Grandes	36	22	8	36.4 %
Medianas T1	374	33	12	36.4 %
Pequeñas	1,955	79	28	35.4 %
Micro	85,910	74	26	35.1 %
Inclusión Forzosa	–	293	0	0.0 %
Total	87,275	501	74	14.8 %

Observación Crítica

La no-respuesta total en inclusión forzosa requiere tratamiento especial mediante imputación basada en registros administrativos

Calibración a Totales Conocidos

Variables de Calibración

- 1 **Facturación total por estrato:** $T_{x,h}$ desde ATM
- 2 **Número de empresas por estrato:** N_h
- 3 **Facturación por sector CLANAE:** Opcional, según convergencia

Restricciones de Calibración

Para cada estrato h :

$$\sum_{i \in r_h} w_i x_i = T_{x,h} \quad (\text{facturación total conocida})$$

$$\sum_{i \in r_h} w_i = N_h \quad (\text{número de empresas})$$

Algoritmo de Implementación

Algorithm 1 Estimador Ratio-GREG con IPW

- 1: **Input:** Datos respondientes (y_i, x_i, R_i) , totales $T_{x,h}$
 - 2: Estimar modelo logit: $\hat{\phi}_i = \mathbb{P}(R_i = 1 | \mathbf{x}_i)$
 - 3: Calcular pesos IPW: $d_i^* = d_i / \hat{\phi}_i$
 - 4: **for** cada estrato h **do**
 - 5: Calibrar pesos: $w_i = d_i^* g_i(\lambda_h)$ tal que $\sum w_i x_i = T_{x,h}$
 - 6: Estimar razón: $\hat{R}_h = \sum w_i y_i / \sum w_i x_i$
 - 7: Calcular VAB: $\hat{Y}_h = \hat{R}_h \times T_{x,h}$
 - 8: **end for**
 - 9: **Output:** PBG total = $\sum_h \hat{Y}_h$
-

Estimación de Varianza

Método Bootstrap Calibrado

- 1 Generar $B = 1000$ muestras bootstrap de unidades respondientes
- 2 Para cada muestra b :
 - Reestimar modelo de propensión
 - Recalcular pesos IPW
 - Recalibrar a totales conocidos
 - Obtener estimación $\hat{Y}^{(b)}$
- 3 Varianza bootstrap: $\hat{V}_{boot} = \frac{1}{B-1} \sum_{b=1}^B (\hat{Y}^{(b)} - \bar{Y}^*)^2$

Intervalos de Confianza

$$IC_{95\%} = \left[\hat{Y} \pm 1,96 \sqrt{\hat{V}_{boot}} \right]$$

Diagnósticos del Modelo de Propensión

Evaluación del Ajuste

- Test Hosmer-Lemeshow
- Curva ROC y AUC
- Análisis de residuos
- Balance de covariables

Tratamiento de Pesos Extremos

- Identificar: $w_i > 5 \times \text{mediana}(w)$
- Opciones:
 - Truncamiento
 - Winsorización
 - Suavizado

Validación de Supuestos

- 1 **MAR:** Análisis de sensibilidad
- 2 **Linealidad:** Gráficos de dispersión parcial
- 3 **Homogeneidad:** Test de razones por estrato
- 4 **Correlación:** $\rho(x_i, y_i) > 0,7$

Análisis de Sensibilidad

Escenarios a Evaluar

- 1 **MNAR débil:** $\mathbb{P}(R|Y, X) = \mathbb{P}(R|X) \times (1 + \delta Y)$
- 2 Diferentes especificaciones del modelo logit
- 3 Métodos alternativos de calibración
- 4 Estratificaciones alternativas

Comparaciones Metodológicas

Método	Sesgo Relativo	RECM
Sin ajuste	Alto	Alto
Solo IPW	Medio	Alto
Solo Calibración	Medio	Medio
IPW + Calibración	Bajo	Bajo

Ventajas de la Metodología Propuesta

Beneficios Técnicos

- **Doble robustez:** Protección contra especificación incorrecta
- **Coherencia externa:** Consistencia con totales ATM
- **Reducción de sesgo:** 20-45 % según evidencia internacional
- **Mejora en precisión:** Reducción de varianza 25-40 %

Beneficios Operativos

- Aprovecha información administrativa existente
- Metodología estándar en institutos estadísticos internacionales
- Implementable con software estadístico disponible (R/Python)
- Escalable a futuros operativos

Recomendaciones para la Implementación

Acciones Inmediatas

- 1 Completar análisis exploratorio de patrones de respuesta
- 2 Validar correlación facturación-VAB en muestra respondente
- 3 Especificar y estimar modelo de propensión
- 4 Implementar calibración a totales ATM

Mejoras Futuras

- Incorporar más variables auxiliares (empleo SIPA, IIBB)
- Desarrollar modelos de imputación para inclusión forzosa
- Establecer sistema de monitoreo continuo
- Documentar exhaustivamente el proceso

Datos precisos para el futuro estratégico provincial

Referencias Bibliográficas

- ① Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376-382.
- ② Särndal, C. E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.
- ③ Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse. *Survey Methodology*, 32(2), 133-142.
- ④ Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance? *Survey Methodology*, 31(2), 161-168.
- ⑤ Haziza, D., & Beaumont, J. F. (2017). Construction of weights in surveys. *International Statistical Review*, 85(1), 79-97.
- ⑥ EUROSTAT. (2021). *Methodology of modern business statistics*. Publications Office of the European Union.

Detalles Técnicos: Función de Calibración

Derivación de los Pesos GREG

La función de distancia chi-cuadrado:

$$G(u) = \frac{1}{2}u^2$$

lleva a la función lagrangiana:

$$\mathcal{L} = \sum_{i \in r} \frac{(w_i - d_i^*)^2}{2d_i^*} - \lambda' \left(\sum_{i \in r} w_i \mathbf{x}_i - \mathbf{T}_x \right)$$

Condición de primer orden:

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{w_i - d_i^*}{d_i^*} - \lambda' \mathbf{x}_i = 0$$

Solución:

$$w_i = d_i^* (1 + \lambda' \mathbf{x}_i)$$

Implementación en R

```
# Ejemplo de código R para implementación
library(survey)
library(sampling)

# 1. Modelo de propensión
modelo_prop <- glm(respuesta ~ log(facturacion) + estrato +
                    sector + departamento,
                    family = binomial, data = muestra)

# 2. Pesos IPW
muestra$prop_est <- predict(modelo_prop, type = "response")
muestra$peso_ipw <- muestra$peso_diseno / muestra$prop_est

# 3. Calibración
diseño_ipw <- svydesign(ids = ~1, weights = ~peso_ipw,
                      data = respondentes)

diseño_cal <- calibrate(diseño_ipw,
                       formula = ~facturacion + estrato,
                       population = totales_poblacionales)
```


SERIE PRODUCTO BRUTO GEOGRÁFICO DE MISIONES – N° 2



25 de Mayo 1460 2° y 3° piso

Tel: 0376 - 4447014 / 4447018

Web: www.ipec.misiones.gov.ar

Email: ipec@misiones.gov.ar

CP 3300 Posadas - Misiones



Sante Fé 343

Tel: 0376-4447637

Web: www.trabajo.misiones.gob.ar

Email: ministerio@trabajo.misiones.gob.ar

CP 3300 Posadas - Misiones