

Estadística Inferencial sobre el analfabetismo en el Perú

Alumnos: Dante Prado Ñaupá , Luis Felix Rojas

Departamento de Ingeniería Económica -UNI

Prof. J. Cerda-Hernández[†]

[†]jcerdah@uni.edu.pe

Asignatura: Estadística Aplicada II

24 de junio de 2019

El presente trabajo utiliza la inferencia estadística para hacer un análisis de la tasa de analfabetismo en algunas regiones del Perú. De acuerdo al último censo realizado por el INEI, estos resultados muestran que en algunos departamentos esta variable aumentó 2 o 3 puntos porcentuales, sin ser tomada en cuenta por el gobierno y la prensa dominante nacional, para mejorar las políticas educativas y sociales. En base a estos resultados, el presente trabajo se propone hacer un análisis inferencial de la tasa de analfabetismo en Puno y Cuzco usando las herramientas vistas en clase, comparando los datos de los dos últimos censos nacionales, 2007 y 2017, mostrando evidencia estadística de dicho aumento.

Palabras Claves: intervalos de confianza, prueba de hipótesis, tasa de analfabetismo, prueba de independencia.

This work is

Keywords: confidence interval, hypothesis test, illiteracy rate, independence test.

1. Introducción

El *analfabetismo* para algunos es simplemente la incapacidad de leer y escribir que se debe generalmente a la falta de enseñanza de las mismas capacidades. Avala esta definición el hecho de que en los países que tienen una escolarización obligatoria el analfabetismo es minoritario.

Según las Naciones Unidas, una persona analfabeta es aquella que no puede ni leer ni escribir un breve y simple mensaje relacionado con su vida diaria. Aunque la educación básica (primaria y secundaria) cada vez llega a más lugares, muchos jóvenes, por diversas razones, no asisten a la escuela y, por lo tanto, algunos de ellos no saben leer ni escribir.

Según la UNESCO, el nivel de iletrados en el Perú llega a los dos millones 211 mil personas aproximadamente, es decir, un 12,3% de la población, de los cuales el 74,1% son mujeres. Este fenómeno muy generalizado en nuestro país, trae

consigo, lo que se mencionó anteriormente, pobreza y exclusión social.

En el caso de la región de Cusco todavía se encuentra entre las regiones con más analfabetismo de jóvenes de 15 a más años en el Perú con 14,78% de acuerdo al INEI, según el último censo 2017.

Del mismo modo, las cifras revelan que los jóvenes con menor probabilidad de acceder a la educación superior se encuentran en el sector de extrema pobreza, provienen de escuelas públicas, áreas rurales o sus padres no lograron culminar estudios primarios o secundarios.

En el caso de Puno, el analfabetismo es un problema latente a nivel del departamento, y se estima que al menos el 12,9% de la población padezca de este problema, refirió el gerente de desarrollo social del Gobierno Regional de Puno (GRP). Hay un problema muy grave, pues durante el gobierno del APRA sus indicadores afirmaban que el Perú en general estaba libre del

analfabetismo, las cifras son optimistas pues revelaban que sólo el 7 % de la población tiene este problema, pero es necesario analizar esta variable con detalle. En ese sentido, el presente trabajo pretende hacer una comparación de las cifras mostradas en los dos últimos censos sobre la tasa de analfabetismo en dichas regiones (censo del 2007 y del 2017), y haciendo un análisis estadístico para ver si realmente dicha tasa disminuyó en esos 10 años transcurridos. El presente trabajo hace un estudio exclusivamente de los departamentos de Puno y Cusco, dejando para trabajos futuros el análisis global de dicha variable.

2. Descripción del problema

El presente trabajo se enmarca en el proceso de aprendizaje asociado a un segundo curso de estadística aplicada de la carrera de ingeniería económica, por lo que el siguiente trabajo se define como un ejercicio empírico para afianzar los conocimientos aprendidos en clase. La estadística inferencial sirve para analizar datos (variables cuantitativas y cualitativas) y tomar decisiones para mejorar las políticas públicas. En ese sentido, el presente trabajo pretende hacer un análisis estadístico de la tasa de analfabetismo en dos departamentos del Perú: Puno y Cusco.

En la actualidad, el Ministerio de Educación (Minedu) viene desarrollando una serie de estrategias y acciones descentralizadas con el objetivo de disminuir la tasa de analfabetismo de 5,9 % a 3,8 % de la población para el 2021, Año del Bicentenario. Uno de ellos fue el Programa nacional de alfabetización (2002-2012), y actualmente políticas educativas que están incluidas en plan a miras al bicentenario (2011-2021). Dichos programas son detallados a continuación:

- Programa nacional de alfabetización (2002-2012)¹: Programa Estratégico, del Ministerio de Educación que se orienta a definir políticas para afrontar el analfabetismo en nuestro país. Resolver este reto social y ético plantea la necesidad de articular esfuerzos desde el Estado y desde la

Sociedad Civil, a través de planes nacionales, regionales y locales que respondan a este imperativo atendiendo a las particularidades socioculturales y lingüísticas de cada lugar.

- Políticas educativas²: estas políticas se enfocan en el acceso a la educación pues esta es la base para el desarrollo humano, y constituye el fundamento de la superación individual para alcanzar el bienestar de las personas. En tal sentido, la erradicación del analfabetismo contribuirá a la inclusión social, la productividad económica y el ejercicio de la ciudadanía. El Plan Bicentenario plantea que una población de 33 millones (para el 2021) de peruanos sin pobreza extrema, desempleo, desnutrición, analfabetismo ni mortalidad infantil. Aquí se proyecta que los peruanos sepan leer y escribir, y tengan una educación básica (terminar la escuela 5to de secundaria)

En base a estos dos programas estratégicos del MINEDU, el presente trabajo pretende comprobar estadísticamente si dichos planes estratégicos para disminuir el analfabetismo fueron efectivos en los departamentos de Puno y Cusco, usando la información brindada por el INEI en los censos del 2007 y del 2017. La pregunta que queremos resolver en este trabajo es

¿Los programas de alfabetización ejecutados por el gobierno central han reducido significativamente el analfabetismo en los departamentos de Puno y Cuzco?

3. Objetivos

3.1. General

El objetivo principal del presente trabajo es determinar el impacto de los programas de alfabetización del MINEDU en los departamentos de Puno y Cuzco, usando como información los datos de los censos del 2007 y 2017 del INEI.

¹Fuente: MINEDU

²Fuente: MINEDU

3.2. Específicos

- Encontrar los datos estadísticos adecuados (media poblacional y muestra, varianza poblacional y muestral, covarianza) para comparar los datos del 2007 y del 2017.
- Construir pruebas de hipótesis para analizar los datos. Para ello trabajaremos primero con una prueba de hipótesis para la tasa de analfabetismo en cada departamento, y después compararemos ambos resultados a través de pruebas de hipótesis para diferencia de proporciones y medias.

4. Justificación del trabajo

El presente trabajo pretende informar y dar un panorama actual del analfabetismo en el Perú, como problema de la mala implementación de políticas educativas y sociales en los países emergentes. La educación sigue siendo un privilegio de las poblaciones con más recursos, ya sea en el Perú o en el resto del mundo. En el caso del Perú, el acceso a una educación básica de calidad sigue siendo uno de los principales problemas del país, sobre todo en las zonas rurales donde el estado no está presente. Problemas de infraestructura y falta de inversión en políticas sociales y educativas a largo plazo son los principales factores que pueden ser consecuencia del aumento del analfabetismo.

El analfabetismo genera brechas sociales. El desarrollo y crecimiento del Perú sería más eficiente y menos desigual si no existiera una gran cantidad de individuos analfabetos, puesto que esta situación retrasa nuestro desarrollo como sociedad, y genera desigualdad y conflictos sociales afectando principalmente el desarrollo del país. Desde el punto de vista socio-político, el analfabeto no puede entender cuáles son sus derechos, debido a lo cual es marginado por la sociedad y engañado por los políticos de turno. Esto también, es motivo para aprovecharse de su ignorancia y explotarlo laboralmente. Si bien los analfabetos ya viven lejos de una sociedad “civilizada”, este problema hace que se mantengan como tal, cerrando las puertas a diversas oportunidades y a una mejor calidad de vida para él y su familia.

El analfabetismo genera en el país consecuencias que se reflejan en tres aspectos fundamentales: social, político y económico. Por ello, analizar y monitorear la dinámica de esta variable es de vital importancia para ver la efectividad de los programas sociales del MINEDU, y reducir la desigualdad entre los peruanos. Desde nuestro curso, podemos hacer un primer análisis de la tasa de analfabetismo para concientizar a la población de este gran problema que aun aqueja a nuestra sociedad, y que la gran prensa nacional oculta, o no da la debida importancia.

5. Marco Teórico

La manera en que analizaremos los datos de los departamentos seleccionados es la siguiente:

En cada censo, se considera una muestra aleatoria simple

censo 2007: X_1, X_2, \dots, X_n

censo 2017: Y_1, Y_2, \dots, Y_n

donde X e Y son variables aleatorias de interés, que representarán: (a) el número de analfabetos por distrito del departamento estudiado por cada 1000 habitantes, (b) la proporción de analfabetos en cada distrito que forma parte de la muestra. Para poder usar a cada distrito como unidad de estudio, supondremos que la media y la proporción de analfabetos en cada distrito es independiente del total de habitantes, es decir hay un nivel alto de homogeneidad en la variable de interés.

Primero, para hacer un estudio individual para cada departamento usaremos el estimador de la proporción $\hat{p} = \frac{X}{n}$, donde X es el número de casos favorables, y n el tamaño de la muestra. Por el Teorema del Límite Central (TLC) tenemos que

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \Rightarrow N(0, 1)$$

donde p es la proporción poblacional del departamento en su respectivo año.

Para analizar la medias de analfabetos por cada 1000 habitantes, usamos el estimador mues-

tral de la media y el TLC,

$$Z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \Rightarrow N(0, 1)$$

donde μ es la media poblacional del departamento en su respectivo año.

De los datos de los dos censos considerados en este trabajo observamos que la proporción de analfabetos y el número de analfabetos por cada 1000 habitantes aumentó, del 2007 para el 2017. Entonces, para contrastar esos resultados previos, observados en los censos del INEI, construiremos intervalos de confianzas y pruebas de hipótesis del tipo

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta > \theta_0$$

con un nivel de significancia de $\alpha = 5\%$, y θ es el parámetro poblacional de interés. Note que planteamos una hipótesis alternativa del tipo $\theta > \theta_0$, siendo θ_0 el valor del parámetro observado en el censo del 2007, porque una estadística simple muestra que la tasa de analfabetismo aumentó en casi dos puntos porcentuales en 10 años, del 2007 al 2017.

Una vez validado realizado el análisis de la media y proporción poblacional mediante la construcción de intervalos de confianza y pruebas podemos responder la pregunta

¿Hubo una mejora significativa en la reducción de analfabetismo, aumentó, o el nivel continúa siendo estadísticamente el mismo como en el 2007?

Otro estudio que podemos realizar usando los datos del INEI, es comparar los parámetros en los dos departamentos, para ver si existe homogeneidad en los datos, y concluir si estadísticamente los valores son iguales.

$$H_0 : \theta_1 = \theta_2 \text{ vs } H_1 : \theta_1 \neq \theta_2$$

Para construir una regla de decisión para esta prueba de hipótesis usamos la distribución muestral para la diferencia de las medias y diferencia de proporciones. En nuestro caso, como los programas implementados por el MINEDU se aplican a todos los departamentos, y además Puno y Cusco son vecinos, podemos incorporar una dependencia entre las distribuciones muestrales

para los dos departamentos. Esto nos lleva a proponer la siguiente versión para la distribución muestral para la diferencia de medias y proporciones

$$\frac{\hat{\theta}_1 - \hat{\theta}_2 - (\theta_1 - \theta_2)}{\sqrt{Var(\hat{\theta}_1 - \hat{\theta}_2)}} \Rightarrow N(0, 1)$$

donde en este caso

$$Var(\hat{\theta}_1 - \hat{\theta}_2) = Var(\hat{\theta}_1) + Var(\hat{\theta}_2) - Cov(\hat{\theta}_1, \hat{\theta}_2)$$

Si n_1 y n_2 son los tamaños de muestra, y suponiendo que las muestras de cada una de las variables aleatorias son independientes se tiene el clásico expresión para la varianza

$$Var(\hat{\theta}_1 - \hat{\theta}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

donde σ_1^2 y σ_2^2 son las varianzas poblacionales. Para ver si existe dependencia o no entre las observaciones en los dos departamentos usamos la prueba de independencia χ^2 (esto para ver si incorporamos o no la covarianza en nuestra prueba y construcción de nuestra regla de decisión).

De manera general, supongamos que se tiene una muestra de n datos bidimensionales de las variables X e Y y que se clasifican en m categorías $\{A_1, A_2, \dots, A_m\}$ para X y k categorías $\{B_1, B_2, \dots, B_k\}$ para Y , en la siguiente tabla de frecuencia conjunta:

$X \setminus Y$	B_1	\dots	B_k	Total
A_1	n_{11}	\dots	n_{1k}	$n_{1\bullet}$
A_2	n_{21}	\dots	n_{2k}	$n_{2\bullet}$
\vdots	\vdots	\ddots	\vdots	\vdots
A_m	n_{m1}	\dots	n_{mk}	$n_{m\bullet}$
Total	$n_{\bullet 1}$	\dots	$n_{\bullet k}$	n

La hipótesis nula para la prueba de independencia es H_0 : las variables X e Y son independientes, y la hipótesis alternativa es H_1 : existe alguna relación entre X e Y . Sobre la hipótesis nula H_0 se espera tener la relación

$$\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j),$$

que es la definición de independencia entre los eventos A_i y B_j . Para verificar la independencia se define el estadístico, introducido por Fisher

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

que asintóticamente tiene una distribución $\chi^2_{(m-1)(k-1)}$, donde m es el número de categorías para la variable X y k es el número de categorías para la variable Y , n_{ij} es la frecuencia observada de la categoría $A_i \cap B_j$ y $e_{ij} = n\mathbb{P}(A_i \cap B_j)$ es la frecuencia esperada de la categoría $A_i \cap B_j$ sobre la hipótesis nula. Sobre la hipótesis nula tenemos

$$e_{ij} = n\mathbb{P}(A_i \cap B_j) = n\mathbb{P}(A_i)\mathbb{P}(B_j) = \frac{n_{i\bullet}n_{\bullet j}}{n}$$

Una vez identificada la relación que tienen las muestras (dependencia o independencia) se procede a construir la regla de decisión con un nivel de significancia α .

6. Descripción de los datos

En este trabajo seleccionamos los departamentos de Puno y Cusco para realizar nuestro análisis. A continuación mostramos los datos tomados de los censos de 2007 y 2017.

Puno: El Departamento de Puno tenía 109 distritos en el 2007 y 110 distritos en el 2017. La población alcanzó los 865421 y 1121728 habitantes en total, con 106871 y 156366 analfabetos, respectivamente, que representan el 12,35 % y 13,94 % de la población de cada departamento, en los dos últimos censos realizados por el INEI.

	Censo 2007	Censo 2017
Población	865421	1121728
Analfabetos	106871 (12.35 %)	156366 (13.94 %)
Analf/1000 hab.	123.49	139.4

Observando los datos que muestran los dos últimos censos observamos que hubo un ligero aumento en el nivel de analfabetismo en el departamento de Cusco.

Cusco: El Departamento de Cusco tenía 108 distritos en el 2007 y 112 distritos en el 2017. La población alcanzó los 768708 y 1147188 habitantes en total, con 107050 y 169524 analfabetos, respectivamente, que representan el 13,93 % y 14,78 % de la población de cada departamento, en los dos últimos censos realizados por el INEI.

	Censo 2007	Censo 2017
Población	768708	1147188
Analfabetos	107050 (13.93 %)	169524 (14.78 %)
Analf/1000 hab.	139.26	147.77

Observando los primeros datos que muestran los dos últimos censos observamos que hubo un ligero aumento en el nivel de analfabetismo en el departamento de Puno.

7. Tratamiento de los datos

Los datos fueron tomados del Instituto de Estadística e Informática INEI <https://www.inei.gob.pe/bases-de-datos/>. Los censos que fueron considerados son los censos nacionales del 2007 y del 2017.

Analfabetismo en Cusco: Según los censos considerados la tasa de analfabetismo son 13,93 % y 14,78 % para el año 2007 y 2017, respectivamente.

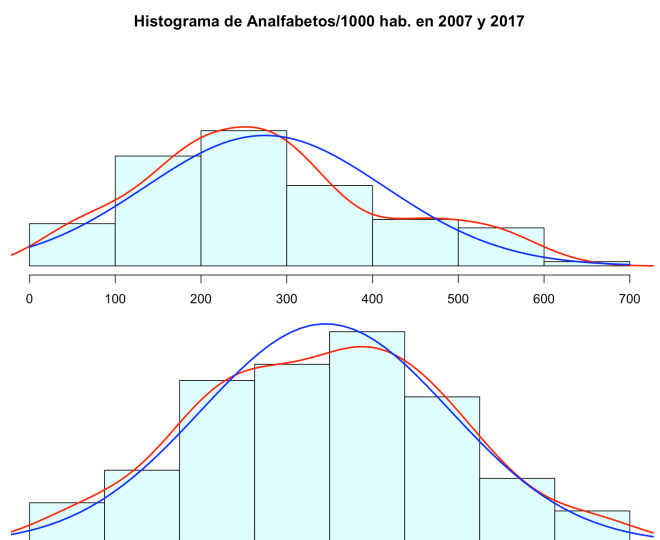


Figura 1. La figura muestra la distribución del número de analfabetos por cada 1000 habitantes en el Departamento de Cusco, en los dos últimos censos nacionales.

Según los censos el departamento cuenta con 123 analfabetos por cada 1000 habitantes en el 2007, y con 139 analfabetos por cada 1000 habitantes en el 2017, con una media y varianza de $\bar{x}_{2007} = 274,6$ y $S^2_{2007} = 139,54$ para el 2007, y $\bar{x}_{2017} = 247,3$ y $S^2_{2017} = 82,86$ para el 2017.

Analfabetismo en Puno: Según los censos considerados en el presente trabajo la tasa de analfabetismo son 12,35 % y 13,94 %, respectivamente.

Histograma de Analfabetos/1000 hab. en 2007 y 2017

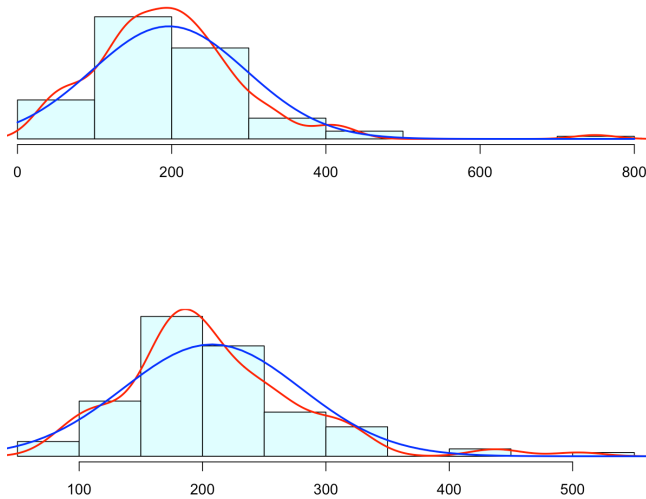


Figura 2. La figura muestra la distribución del número de analfabetos por cada 1000 habitantes en el Departamento de Puno, en los dos últimos censos nacionales.

Las Figuras 1 y 2 muestran una ligera translación a la derecha de la media, mostrando un posible aumento en la media de analfabetos por cada 1000 habitantes. Para contrastar los resultados de los censos construimos intervalos de confianza y pruebas de hipótesis adecuadas. La siguiente tabla muestra estos resultados, para un nivel de significancia de $\alpha = 0,05$.

	$\hat{\theta}$	L	S
$\bar{x}_{pun,2007}$	196.49	177.51	215.47
$\bar{x}_{pun,2017}$	207.77	194.12	221.422
$\bar{x}_{cuz,2007}$	991.2	842.92	1139.78
$\bar{x}_{cuz,2017}$	1513.6	1198.8	1828.4
$\hat{p}_{pun,2007}$	12.35	9.25	15.45
$\hat{p}_{pun,2017}$	13.94	7.35	20.5
$\hat{p}_{cuz,2007}$	13.93	7.35	20.5
$\hat{p}_{cuz,2017}$	14.78	8.03	21.52

Para ver si estadísticamente el parámetro de interés aumenta o es igual con un nivel de significancia de 5%, construimos una prueba de hipótesis de la forma

$$H_0 : \theta_1 = \theta_2 \text{ vs } H_1 : \theta_1 > \theta_2$$

donde θ_1 representa el valor del parámetro en el 2007, y θ_2 el valor del parámetro en el 2017.

Para el caso de la media en Cusco tenemos la prueba de hipótesis $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_2 - \mu_1 > 0$. En este caso la diferencia de medias observadas fue de 522.4 analfabetos por cada 1000 habitantes dando un valor

$t(\text{student})=0.030345$ que no pertenece a la región crítica $RC=[1.676;t]$, obligando a aceptar la hipótesis nula, es decir, la media de analfabetos por cada 1000 habitantes en el Departamento de Cusco no cambió significativamente del ao 2007 al ao 2017.

Para el caso de la media en Puno tenemos la prueba de hipótesis $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_2 - \mu_1 > 0$. En este caso la diferencia de medias observadas fue de 11.28 analfabetos por cada 1000 habitantes dando un el valor $t(\text{student})=0.091413$ que no pertenece a la región crítica $RC = [1,676 < t]$ obligando a no rechazar la hipótesis nula, es decir, la media de analfabetos por cada 1000 habitantes en el 2017 en el Departamento de Puno se mantiene al mismo nivel de 2007.

Para el caso de la proporción de analfabetos en Cusco tenemos la prueba de hipótesis $H_0 : \hat{p}_1 - \hat{p}_2 = 0$ contra $H_1 : \hat{p}_1 - \hat{p}_2 > 0$. En este caso la diferencia de proporciones observadas fue de 0,85 %, y su probabilidad $P=P(t_i 1.179)$ que pertenece a la región crítica, $RC = P[1,676 < t]$ obligando rechazar la hipótesis nula, es decir, la proporción de analfabetos en el Departamento de Cusco aumentó de 2007 para el 2017.

Para el caso de la proporción de analfabetos en Puno tenemos la prueba de hipótesis $H_0 : \hat{p}_1 - \hat{p}_2 = 0$ contra $H_1 : \hat{p}_1 - \hat{p}_2 > 0$. En este caso la diferencia de proporciones observadas fue de 1,59 %, y su probabilidad $P=P(t_i 0.58)$ que no pertenece a la región crítica, $RC = P[1,676 < t]$ obligando no rechazar la hipótesis nula, es decir, la proporción de analfabetos en el 2017 en el Departamento de Cusco es igual a la del 2007, estadísticamente con un nivel de significancia de 5%.

8. Conclusiones y posibles extensiones

Con los resultados hallados mediante la prueba de hipótesis de la diferencia de medias y proporciones con varianza poblacional conocidas aplicada a los departamentos de Cusco y Puno concluimos que para ambos caso no existe una mejora significativa en reducir el analfabetismo

en dichos departamentos, ya que mediante nuestra prueba de hipótesis para ambos casos las medias poblacionales son estadísticamente iguales (con los datos hallados en el CENSO 2007 y el CENSO 2017). Incluso en algunos casos nos salió que hubo un aumento para los parámetros estudiados, como la media para el Departamento de Cusco y la proporción para el Departamento de Puno, pero en ninguno caso salió que hubo evi-

dencia estadística de una disminución. Por ello se concluye con un nivel de significancia de 5% que el programa nacional de alfabetización no estará cumpliendo con su objetivo de erradicar el analfabetismo proyectado para el 2021 (Plan Bicentenario).

Para tener más evidencia, como trabajo futuro se hará un contraste en todos los departamentos usando modelos lineales.

-
1. Montgomery, D. and Runger G. *Applied Statistics and Probability for Engineers*. Hoboken, NJ: Wiley, (2007).
 2. Instituto de Estadística e Informática - INEI.

Censo Nacional 2007.

3. Instituto de Estadística e Informática - INEI. Censo Nacional 2017.