Familia y calidad de vida en las provincias argentinas

Sobre datos del censo 2010

Especialización en Producción y Análisis de Información para las Políticas Publicas Seminario: Análisis de Datos Cuantitativos

Autor: Andrés Vázquez andres@data99.com.ar

Julio de 2014 Córdoba, Argentina

Introducción

El presente estudio está basado en datos del censo argentino 2010 realizado por el INDEC¹ en octubre de ese año. Se utilizarán 11 variables para los 52.383 radios censales usados por el INDEC para cubrir toda la superficie del territorio argentino. Los análisis serán siempre por provincia permitiendo comparar las realidades de cada distrito. Los datos de base están en formato CSV y pueden descargarse².

El término *familia* en este estudio se refiere simplemente a la existencia de cónyuge e hijos en los hogares. No se corresponde con ninguna definición formal y solo se limita a los alcances de los datos usados. Las comparaciones estarán basadas en relacionar indicadores de *familia* y calidad de vida pensando en entender si existe relación entre estas. El término *calidad de vida* en este informe se mide con las variables: a) existencia de baño en el hogar, b) ausencia de NBI (necesidades básicas insatisfechas), c) titulo de propiedad de la vivienda, d) calidad de los materiales de construcción, e) vivienda tipo "casa" (por sobre departamento, inquilinato, etc), f) existencia de computadora en el hogar y g) asistencia de los integrantes del hogar a alguna institución educativa.

La **hipótesis** a validar es que los indicadores de *familia* influyen positivamente en la *calidad de vida* de los hogares argentinos. Acompañado este dato por las diferencias existentes entre los 24 distritos nacionales (23 provincia y CABA³ que se tratada como una provincia más).

Herramientas a utilizar

Para todos los análisis y gráficos se uso el lenguaje de programación Python con sus librerías SciPy (regresión lineal), NumPy (manejo de números y matrices) y Matplotlib (gráficos). Todos los scripts⁴ utilizados para el proceso se incluyen al final de este trabajo en un anexo técnico. En el mismo anexo se incluyen las descripciones de las variables tomadas de los datos del censo.

Objetivos

- 1. Conocer la proporción de jefes de hogar que conviven con su cónyuge en cada uno de los 24 distritos.
- 2. Calcular la cantidad de hijos promedio por cada hogar en cada provincia.
- 3. Buscar correlación entre la existencia de cónyuges (considerado *familias unidas*) y los indicadores de calidad de vida seleccionados.

^{1 &}lt;a href="http://www.indec.mecon.ar">http://www.indec.mecon.ar

² http://andresvazquez.com.ar/data/censo-2010-familia/ALL-RESULTS.csv

³ Ciudad Autónoma de Buenos Aires.

⁴ Breve porción de código fuente de programación orientado a resolver un problema concreto.

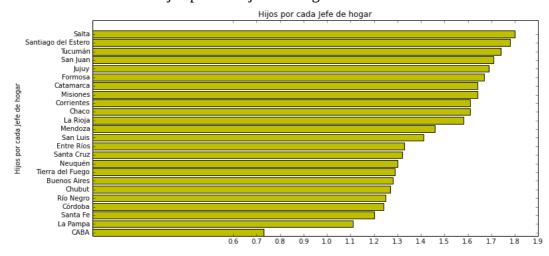
Familias unidas

Se considera familias unidas a aquellas en la que exista un cónyuge en el hogar. La siguiente tabla es el acumulado de personas (total), jefes de hogar y cónyuges por provincia. Esto permite calcular el porcentaje de jefes de hogar que conviven con su pareja.

Provincia	Personas	Jefes de hogar	Cónyuges	% de jefes con pareja
CABA	2.890.151	1.150.732	568.496	49,40
Catamarca	367.828	96.028	53.141	55,34
Jujuy	673.307	174.704	97.910	56,04
Formosa	530.162	140.330	79.862	56,91
La Rioja	333.642	91.112	52.053	57,13
Salta	1.214.441	299.882	173.397	57,82
La Pampa	318.951	107.684	62.796	58,32
Santiago del Estero	874.006	218.073	127.180	58,32
Neuquén	551.266	170.095	100.112	58,86
Santa Cruz	273.964	81.798	48.257	59,00
Corrientes	992.595	267.839	158.256	59,09
Río Negro	638.645	199.206	118.416	59,44
Chubut	509.108	157.215	93.488	59,47
Chaco	1.055.259	288.479	172.090	59,65
Córdoba	3.308.876	1.032.064	617.351	59,82
Tierra del Fuego	127.205	38.977	23.356	59,92
San Luis	432.310	126.943	76.179	60,01
Santa Fe	3.194.537	1.023.908	617.324	60,29
Entre Ríos	1.235.994	375.183	231.085	61,59
Tucumán	1.448.188	368.581	227.389	61,69
Buenos Aires	15.625.084	4.790.063	2.974.276	62,09
Misiones	1.101.593	303.035	193.897	63,99
San Juan	681.055	177.173	114.702	64,74
Mendoza	1.738.929	494.917	323.476	65,36
TOTALES	40.117.096	12.174.021	7.304.489	60,00

Hijos por jefe de hogar

Otro aspecto interesante de la conformación familiar son los hijos en cada familia. Para esto se tomó la cantidad de hijos por cada jefe de hogar.



Este gráfico (junto al anterior) confirman una tendencia de CABA como distrito que incluye la menor cantidad de familias unidas e hijos por cada hogar.

Familia y calidad de vida

La variable *cónyuges* (que interpretamos como *familias unidas*) muestra un coeficiente de correlación muy alto en relación a **todas** las variables elegidas para *calidad de vida*. Esto no es un indicador claro de que las familias unidas expliquen estos indicadores positivos de calidad de vida ya que, además, estos indicadores de *calidad de vida* elegidos, tienen también altos índices de correlación para los hogares en general, independientemente de la existencia de cónyuges.

La comparación adicional elegida es la de *jefes de hogar* (variable P__01_01, elegida como representación del total de hogares en general) contra los mismo indicadores de *calidad de vida*. De esta forma cuando lo indicadores de correlación (R²) entre *cónyuges* y *calidad de vida* sean superiores a *jefes de hogar* versus *calidad de vida* podremos entender que estos ejercen una influencia positiva.

Familia y asistencia a instituciones educativas

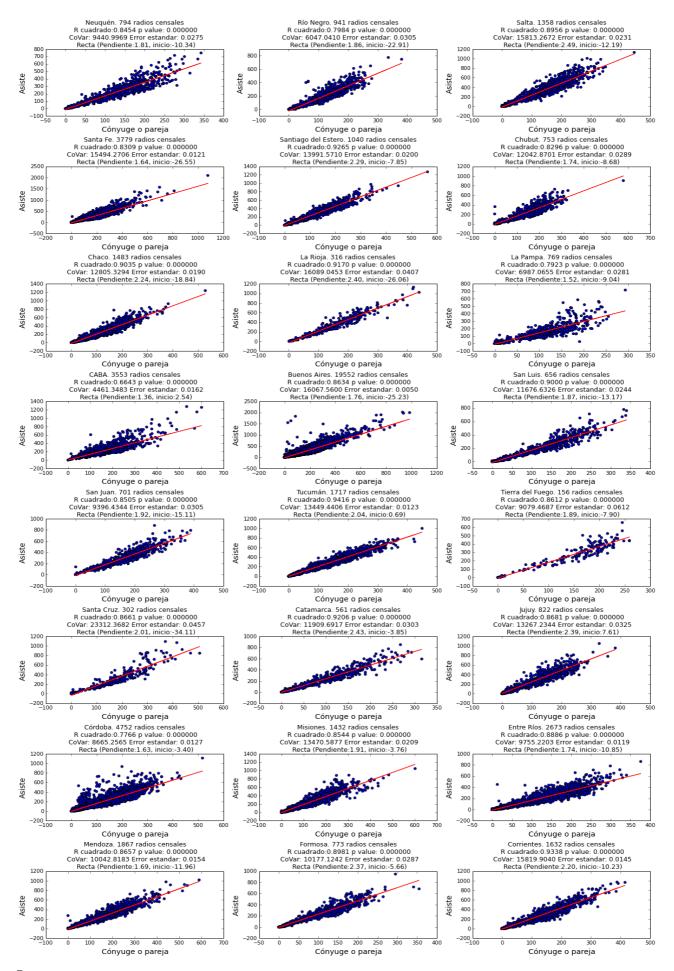
Para esta serie de análisis de **regresión lineal** se utilizó la función linregress⁵ de la librería SciPy, que permite obtener simultáneamente el coeficiente de correlación **r** (que se usara al cuadrado), el valor **p** (que en todos los casos nos permite rechazar la hipótesis nula) y los valores necesarios para la graficar la recta de regresión.

Los siguientes 24 diagramas de dispersión (uno por distrito) analizan la correlación entre la cantidad de cónyuges (variable en la base de datos P_01_02) y personas que asisten a la escuela (variable P_08_01). Se usa la cantidad de cónyuges para tratar de conocer si el hecho de que las familias estén unidas pueda afectar positivamente la asistencia a la escuela de cualquiera de los integrantes del hogar⁶.

La hipótesis (H_0) de que cantidad de cónyuges y personas que asisten a la escuela son variables independientes puede rechazarse según los **p-values** obtenidos para cada distrito.

⁵ http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html

⁶ No discriminado por tipo de integrantes que asiste a institución educativa, los datos usados tiene esa limitación.



Familias unidas y otros indicadores de calidad de vida

Otras variables analizadas muestran fuertes correlaciones (R²) con la cantidad de cónyuges pero **en ningún caso son superiores** a las tomadas en comparación con el total de hogares (según la variable *jefes de hogar*). Las variables:

- 1. Existencia de baño en el hogar
- 2. Ausencia de NBI (necesidades básicas insatisfechas)
- 3. Titulo de propiedad de la vivienda
- 4. Calidad de los materiales de construcción
- 5. Tipo de vivienda "casa" (por sobre departamento, rancho, etc)
- 6. Existencia de computadora en el hogar

Tienen fuerte correlación con la cantidad de cónyuges pero esta correlación es **menor** a la correlación que existe con el total de hogares (independientemente de la cantidad de cónyuges). Es por esto que no podemos atribuir esta *calidad de vida* a la existencia de *familias unidas* aún cuando los indicadores lo sugieran.

Conclusiones

En este estudio no es posible afirmar que la existencia de *familias unidas* sea un indicador que incida directamente en la calidad de vida de los hogares argentinos. En cambio, sí es posible afirmar que la educación (entendida como asistencia a los institutos educativos) esta relacionada con la convivencia en pareja de **los jefes** del hogar.

La Ciudad Autónoma de Buenos Aires aparece claramente como la menos propensa a contener familias en los términos de cónyuges por jefe de hogar e hijos en cada grupo familiar. Es notoriamente inferior a todos los demás distritos. En el estudio de correlación entre cónyuges y asistentes a las escuelas tiene además un valor inferior comparado con todas las provincias. Claramente, CABA no es el distrito de las familias. Por el contrario no es posible afirmar que existan provincias con una tendencia opuesta o claramente favorable a la familia (en los términos entendidos en este estudio).

Por otro lado, sobre distritos norte argentino es posible notar las siguientes tendencias:

- De los once primeros distritos con mayor cantidad de hijos por jefe de hogar, diez pertenecen al norte argentino.
- En el caso de la correlación entre existencia de cónyuges y asistencia escolar los cinco distritos con mayor vinculación pertenecen también al norte argentino.
- Los cinco distritos con menor porcentaje de cónyuges por jefe de hogar (después de CABA) pertenecen también al norte argentino.

Anexo técnico

Variables disponibles en los datos usados

Están marcadas en rojo aquellas variables que si fueron usadas

Estan marcadas en rojo aquenas variables que si r	deroii dodddo
H_05 Material predominante de los pisos	H_INDHAC_01 Hasta 0.50 personas por cuarto
H_05_01 Cerámica, baldosa, mosaico, mármol, madera o alfombrado	H_INDHAC_02 0.51 - 0.99 personas por cuarto
H_05_02 Cemento o ladrillo fijo	H_INDHAC_03 1.00 - 1.49 personas por cuarto
H_05_03 Tierra o ladrillo suelto	H_INDHAC_04 1.50 - 1.99 personas por cuarto
H_05_04 Otro	H_INDHAC_05 2.00 - 3.00 personas por cuarto
H_06 Material de la cubierta exterior del techo	H_INDHAC_06 Más de 3.00 personas por cuarto
H06_01 Cubierta asfáltica o membrana	H_PROP Posesión del inmueble
H06_02 Baldosa o losa (sin cubierta)	HPROP_01 Propietario de la vivienda y del terreno
H06_03 Pizarra o teja	HPROP_02 Propietario sólo de la vivienda
H06_04 Chapa de metal (sin cubierta)	H_PROP_03 Inquilino
H_06_05 Chapa fibrocemento o plástico	H_PROP_04 Ocupante por préstamo
H_06_06 Chapa de cartón	H_PROP_05 Ocupante por trabajo
H06_07 Caña, palma, tabla o paja con o sin barro	H_PROP_06 Otra situación
H_06_08 Otro	H_TOTPERS Total de personas en el hogar
H_07 Revestimiento interior o cielorraso del techo	HTOTPERS_01 1 persona en el hogar
H07_01 Con revestimiento interior o cielorraso del techo	HTOTPERS_02 2 personas en el hogar
H07_02 Sin revestimiento interior o cielorraso del techo	HTOTPERS_03 3 personas en el hogar
H_08 Tenencia de agua	HTOTPERS_04 4 personas en el hogar
H08_01 Por cañeria dentro de la vivienda	HTOTPERS_05 5 personas en el hogar
H08_02 Fuera de la vivienda pero dentro del terreno	H_TOTPERS_06 6 personas en el hogar
H08_03 Fuera del terreno	H_TOTPERS_07 7 personas en el hogar
H_09 Procedencia del agua para beber y cocinar	H_TOTPERS_08 8 o más personas en el hogar
H09_01 Red pública	P_01 Relación o parentesco con el jefe(a) del hogar
H_09_02 Perforación con bomba a motor	P01_01 Jefe(a)
H09_03 Perforación con bomba manual	P01_02 Cónyuge o pareja
H09_04 Pozo	P01_03 Hijo(a) / Hijastro(a)
H09_05 Transporte por cisterna	P01_04 Yerno / Nuera
H09_06 Agua de lluvia, río, canal, arroyo o acequia	P01_05 Nieto(a)
H_10 HOGAR Tiene baño o letrina	P01_06 Padre / Madre / Suegro(a)
H 10 01 HOGAR SI Tiene baño o letrina	P 01 07 Otros familiares
H10_02 HOGAR NO Tiene baño o letrina	P01_08 Otros no familiares
H_11 Tiene botón, cadena, mochila para limpieza del inodoro	P_01_09 Servicio doméstico y sus familiares
H_11_01 Si tiene cadena y mochila en el baño	P 02 Sexo
H_11_02 No tiene cadena y mochila en el baño	P_02_01 Hombre
H_12 Desague del inodoro	P_02_02 Mujer
H_12_01 A red publica (cloaca)	P_05 País de nacimiento
H12_02 A cámara séptica y pozo ciego	P05_01 Argentina
H12_03 Sólo a pozo ciego	P05_02 Otros
H12_04 A hoyo, excavación en la tierra, etc.	P_07 Sabe leer y escribir
H_13 Exclusividad del baño para el hogar	P07_01 Sabe leer y escribir
H13_01 De uso exclusivo del hogar	P07_02 No sabe leer y escribir
H_13_02 Baño compartido	P_08 Condicion de asistencia escolar
H_14 Combustible para cocinar	P08_01 Asiste
H 14 01 Gas de red	P08_02 Asistió
H14_02 Gas a granel (zeppelin)	P08_03 Nunca asistio
H_14_03 Gas en tubo	P_09 Nivel educativo que cursa o cursó
H_14_04 Gas en garrafa	P_09_01 Inicial (jardin, preescolar)
H_14_05 Electricidad	P_09_02 Primario
H_14_06 Leña o carbón	P_09_03 EGB
H_14_07 Otro	P09_04 Secundario
H_19A Heladera	P09_05 Polimodal
H19A_01 Con Heladera	P09_06 Superior no universitario
H_19A_02 Sin Heladera	P09_07 Universitario
H_19B Computadora	P09_08 Post universitario
H_19B_01 Con computadora	P09_09 Educacion especial
H19B_02 Sin computadora	P_10 Completo el nivel (P09)
H_19C Teléfono celular	P10_01 Si ompleto el nivel (P09)
H_19C_01 Con teléfono celular	P10_02 No completo el nivel (P09)
H 19C 02 Sin telefono celular	P_12 Utiliza computadora
<u> </u>	
H_19D Teléfono de línea	P_12_01 Si utiliza computadora
H_19D_01 Con teléfono de línea	P_12_02 No utiliza computadora
H_19D_02 Sin teléfono de línea	P_CONDACT Condición de actividad
H_ALGUNBI Al menos un indicador NBI	PCONDACT_01 Ocupado
H_ALGUNBI_01 Sin NBI	
	PCONDACT_02 Desocupado
H_ALGUNBI_02 Al menos un indicador NBI H_INDHAC Indice de hacinamiento	

```
V__01_01 Casa
                                                                       _INMAT_03
                                                                                       Calidad 3
V__01_02 Rancho
                                                                       _INMAT_04
                                                                                       Calidad 4
V__01_03 Casilla
                                                                       TIPVV
                                                                                       Tipo de vivienda agrupado
V__01_04 Departamento
                                                                       _TIPVV_01
                                                                                       Viviendas particulares
                                                                       TIPVV 02
V__01_05 Pieza en inquilinato
                                                                                       Viviendas colectivas
V\_01\_06 Pieza en hotel familiar o pensión
                                                                       TOTHOG
                                                                                       Cantidad de Hogares en la Vivienda
                                                                       TOTHOG_01
V__01_07 Local no construido para habitación
                                                                                       Con un hogar
V__01_08 Vivienda móvil
                                                                   V__TOTHOG_02
                                                                                       Más de un hogar
V__01_09 Persona/s viviendo en la calle
V__02 Condición de ocupación
                                                                       _URP Area Urbano - Rural
                                                                       URP_01
                                                                                       Area Urbano
V__02_01 Con personas presentes
                                                                   V__URP_02
                                                                                       Area Rural agrupado
V__02_02 Con todas las personas temporalmente ausentes
                                                                       _URP_03
                                                                                       Area Rural disperso
V__02_03 En alquiler o venta
                                                                       _V00 Tipo de vivienda colectiva
                                                                       _V00_01
V__02_04 En construcción
                                                                                       Hogar de ancianos
V_02_05 Se usa como comercio, oficina o consultorio
                                                                       _V00_02
                                                                                       Hogar de menores
V__02_06 Se usa para vacaciones, fin de semana u otro uso temporal
                                                                       _V00_03
                                                                                       Colegio internado
V__02_07 Por otra razón
                                                                       V00_04
                                                                                       Campamento/obrador
V_INCALCONS Calidad constructiva de la vivienda
                                                                       V00 05
                                                                                       Hospital
V_INCALCONS_01
                                                                       V00 06
                             Calidad constructiva Satisfactoria
                                                                                       Prisión
V__INCALCONS_02
                                                                       V00_07
                             Calidad constructiva Basico
                                                                                       Cuartel
                                                                                       Hogar de religiosos
V__INCALCONS_03
                             Calidad constructiva Insuficiente
                                                                       _V00_08
V_INCALSERV Calidad de Conexiones a Servicios Basicos
                                                                    V__V00_09
                                                                                       Hotel turístico
V__INCALSERV_01 Calidad de Servicios Basicos Satisfactorio
                                                                       _V00_10
V_INCALSERV_02 Calidad de Servicios Basicos Basico
V__INCALSERV_03 Calidad de Servicios BasicosInsuficiente
                                                                    Personas Cantidad de personas en cada radio censal
V INMAT
                   Calidad de los materiales
                                                                    Hogares Cantidad de hogares en cada radio censal
V__INMAT_01
                                                                    Viviendas Cantidad de viviendas en cada radio censal
                   Calidad 1
V__INMAT_02
                   Calidad 2
```

Scripts usados

Las constantes usadas (provincias y variables) fueron empaquetadas para su reutilización y están disponibles para descarga⁷. Para todas las tareas se incluyen los siguientes encabezados para la carga de las librerías Python a utilizar.

```
import matplotlib.pyplot as plt # importar librería gráfica
import numpy as np # cargar librería numpy para matrices
from scipy import stats # cargar librería SciPy para regresiones lineales.
```

Script para paso de CSV a JSON (uno por provincia)

El formato JSON⁸ es más amigable para el procesamiento por lo que se tomo el fichero original CSV y se construyeron ficheros individuales por provincia en este formato. Se aprovecho para incluir valores para cada variable con el sufijo "_PORC" para tomar además el porcentaje de cada valor dentro del total de cada variable. Por ejemplo P__01_01 es la cantidad de Jefes de hogar en cada radio censal, P__01_02 es la cantidad de cónyuges. Opcionalmente se podrá acceder al porcentaje de Jefes dentro del total de personas y al porcentaje de Cónyuges dentro del mismo total con las variables P__01_01_PORC y P__01_02_PORC. El código es el siguiente:

```
import csv
import json
from pkg import constantes

def csv_to_json(path_file):
    print "READING ..."

for pr, pr_name in constantes.Provincias.iteritems():
```

^{7 &}lt;a href="http://andresvazquez.com.ar/data/censo-2010-familia/pkg/constantes.py">http://andresvazquez.com.ar/data/censo-2010-familia/pkg/constantes.py

^{8 &}lt;a href="http://es.wikipedia.org/wiki/JSON">http://es.wikipedia.org/wiki/JSON

```
f = open(path_file, "r")
        reader = csv.DictReader(f, delimiter="\t")
        print "CREATING %s" % pr_name
        data = []
        for row in reader:
            if int(pr) != int(row["Provincia"]):
                continue
            new_row = {}
            for k, v in row.iteritems():
                try: # quiero los campos numericos como numeros para poder procesar despues
                   new_row[k] = int(v)
                except:
                    new row[k] = v
                else:
                    if k[:3] == "V__":
                        nr = "%s_PORC" % k
                        if int(row["Viviendas"]) > 0:
                            new_row[nr] = round(100 * float(v) / float(row["Viviendas"]),3)
                        else:
                            new_row[nr] = 0
                    if k[:3] == "H___":
                        nr = "%s PORC" % k
                        if int(row["Hogares"]) > 0:
                            new_row[nr] = round(100 * float(v) / float(row["Hogares"]),3)
                        else:
                    if k[:3] == "P__":
nr = "%s_PORC" % k
                        if int(row["Personas"]) > 0:
                            new_row[nr] = round(100 * float(v) / float(row["Personas"]),3)
                        else:
                            new_row[nr] = 0
            data.append(new_row)
        print "SAVING %s (%s)" % (pr_name, len(data))
        g = open("datajson/censo2010.%s.json" % pr , "w")
        g.write(json.dumps(data))
        g.close()
# leer el archivo CSV con datos del censo por provincia y pasarlos a JSON
csv_to_json("ALL-RESULTS.csv")
```

Script "Familias unidas"

```
conyuges = {}
jefes = {}
personas = {}
porc_con_conyuges = {}
for k, v in constantes.Provincias.iteritems():
    conyuges[k] = np.sum(data_series[k]["P__01_02"])
    jefes[k] = np.sum(data_series[k]["P__01_01"])
    personas[k] = np.sum(data_series[k]["Personas"])
    porc_con_conyuges[k] = round(100 * float(conyuges[k]) / float(jefes[k]),2)
    print "%s;%s;%s;%s;%s;%s" % (k, v, str(personas[k]), str(jefes[k]), str(conyuges[k]),
str(porc_con_conyuges[k]))
```

Script "hijos por jefe de hogar"

```
hijos_por_jefe = {}
hijos = {}
for k, v in constantes.Provincias.iteritems():
    hijos[k] = np.sum(data_series[k]["P__01_03"])
    hijos_por_jefe[v.decode("utf-8")] = round(float(hijos[k]) / float(jefes[k]),2)
```

```
sorted_hijos_por_jefe = sorted(hijos_por_jefe, key=hijos_por_jefe.__getitem__)
sort_vals = [hijos_por_jefe[c] for c in sorted_hijos_por_jefe]

fig = plt.figure(figsize=(12, 6))
vals = len(hijos_por_jefe.keys())
plt.barh(range(vals), sort_vals, color='y')
plt.ylabel('Hijos por cada Jefe de hogar')
plt.title('Hijos por cada Jefe de hogar')
width = 0.75
plt.yticks(np.arange(vals) + width /2., np.array(sorted_hijos_por_jefe) )
plt.xticks(np.arange(0.6,2.0,0.1))
# plt.legend( (p1[0], p2[0]), ('Men', 'Women') )
```

Script para "familia y asistencia a la escuela"

Este script es genérico y puede usarse para comparar dos variables cualquiera en los 24 distritos y hacer un diagrama de dispersión junto con la recta de regresión y los valores de regresión correspondientes

```
def compare_vars(title, compare1, compare2):
    fig = plt.figure(figsize=(20, 32)) # cada punto son 80 pixeles. 16 es todo el ancho de mi
pantalla aprox.
    fig.subplots_adjust(hspace=0.75, top=0.90)
    fig.suptitle(title, fontsize=18, fontweight='bold')
    for k, v in constantes.Provincias.iteritems():
        xy = np.vstack((data_series[k][compare1],data_series[k][compare2]))
        cv = np.cov(xy)

covar = "%.4f" % cv[1][0] # desviacion standar
        ax = fig.add_subplot(8, 3, ix)
        ix += 1
        arr1 = np.array(data_series[k][compare1])
        arr2 = np.array(data_series[k][compare2])
        slope, intercept, r_value, p_value, std_err = stats.linregress(arr1, arr2)
        r = "%.4f" % (r_value**2)
p = "%.6f" % p_value
serr = "%.4f" % std_err
s = "%.2f" % slope
        i = "%.2f" % intercept
        ax.set_xlabel(constantes.getFlds(compare1), fontsize=14)
        ax.set_ylabel(constantes.getFlds(compare2), fontsize=14)
        line = slope * arr1 + intercept
ax.set_title('%s. %s radios censales \n R cuadrado:%s p value: %s \n CoVar: %s Error
estandar: %s \n Recta (Pendiente: %s, inicio: %s) ' % (v.decode ("utf-8"), len(arr1), r, p, covar,
serr, s, i))
         # line es [] cuando no hay correlacion
        if len(line) == len(arr1):
            plt.plot(arr1, line, 'r-')
        plt.scatter(arr1, arr2)
    plt.show()
compare_vars("Conyuge vs Asistentes a escuela", "P__01_02", "P__08_01")
```

Índice

Introducción	2
Herramientas a utilizar	2
Objetivos	2
Familias unidas	
Hijos por jefe de hogar	3
Familia y calidad de vida	
Familia y asistencia a instituciones educativas	4
Familias unidas y otros indicadores de calidad de vida	6
Conclusiones	7
Anexo técnico	8
Variables disponibles en los datos usados	8
Scripts usados	9
Script para paso de CSV a JSON (uno por provincia)	9
Script "Familias unidas"	10
Script "hijos por jefe de hogar"	
Script para "familia y asistencia a la escuela"	11