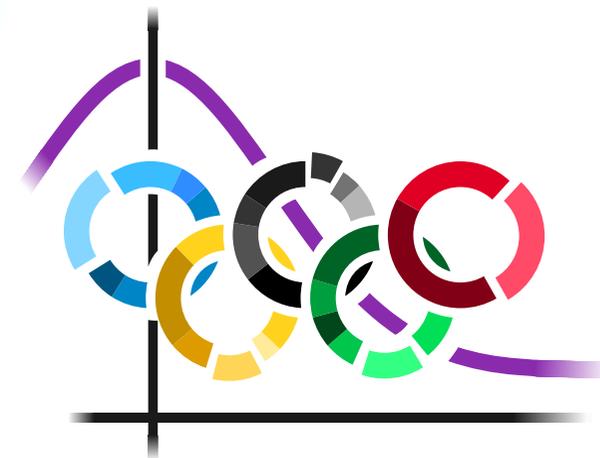


# OLIMPIADA de ESTADÍSTICA 2018



ESTUDIO ESTADÍSTICO  
SOBRE LA

ENCUESTA CONTINUA DE  
**HOGARES**  
2016

**RUBIK**  **BED**

**Equipo RUBIKCUBED (BACHILLERATO)**

INTEGRANTES: Jaime Gómez Benavente  
TUTORA: María Alonso García  
COLEGIO LOURDES (VALLADOLID)

# OBJETIVOS DEL PROYECTO



A partir de los datos de la Encuesta Continua de Hogares, fijamos tres objetivos generales y amplios, que son, como se explica a continuación, el análisis de:

## 1. VARIABLES PRINCIPALES DEL ENCUESTADO

- Variable discreta **Edad**, con **distribución y normalización con campana de Gauss**
- Relación entre **edad media del encuestado y estado civil**
- Relación entre **estudios y actividad laboral** del encuestado
- Estudio del **mercado laboral por sexo**
- Prototipo de **encuestado**, en base a **parámetros** habituales
- **Inferencia estadística** mediante **contraste de hipótesis**

## 2. VARIABLES PRINCIPALES DEL HOGAR

- **Habitantes por hogar**, con **distribución y normalización con campana de Gauss, media, mediana, moda y desv. típica**
- Cartograma con **superficie media en CCAA**, media nacional
- Relación entre **nº de habitantes y superficie en m<sup>2</sup>** de los hogares y viviendas de la Encuesta Continua
- Régimen de **tenencia y ránking** de hogares por m<sup>2</sup>
- Número de **hijos, modelo de hogar y tipos de hogar**

## 3. CORRELACIÓN ENTRE AMBAS VARIABLES

- Relación entre la **España rural y la España urbana**, en materia de **régimen de tenencia y superficie en m<sup>2</sup>**
- Correlación entre **edad, núcleo familiar y habitantes de la vivienda**

## BASE TEÓRICA DEL ESTUDIO TEORÍA DE LA ESTADÍSTICA

Nuestro estudio empleará dos elementos estadísticos que requieren de principio teórico.

### ➤ **TIPIFICACIÓN EN DISTRIBUCIÓN ESTÁNDAR $N(0,1)$**

Para facilitar el análisis, tipificamos las distribuciones de variables discretas en una normal (0,1)

$$P(x = a) \Rightarrow P\left(z = \frac{a - \mu}{\sigma}\right) \triangleright N(0,1)$$

### ➤ **CONTRASTE DE HIPÓTESIS E INFERENCIA ESTADÍSTICA**

Se realiza un contraste unilateral sobre la proporción de hogares, con un nivel de significación de un 1% y el siguiente rango de aceptación

$$RA = \left( \mu - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \mu + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

# MÉTODO EN CINCO PASOS



```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
```

Tomando los datos y categorías proporcionados en los 5000 registros de la **Encuesta Continua de Hogares 2016**, se comienza un **detallado análisis** para escoger los ítems que nos resultarán útiles durante el estudio.

Además, se tienen en cuenta los **diferentes ámbitos de los datos** para agruparlos en **tres objetivos**.

```
''' OLIMPIADA NACIONAL DE ESTADISTICA 2018
ENCUESTA CONTINUA DE HOGARES
PROGRAMA DE ANALISIS DE DATOS. '''
```

Una vez seleccionados los objetivos, se procede a **estudiar las relaciones entre variables** y a seleccionar aquellas que nos resultan curiosas y que, probablemente, **nos lleven a obtener resultados interesantes**.

```
__author__ = 'Jaime Gomez'
__version__ = '1.0'
```

A continuación, se **organizan los datos** (i.e. la base del estudio) en función de los **tres objetivos** elegidos.

```
import csv
```

```
import pandas as pd
import matplotlib.pyplot as plt
import scipy
```

Debido a la **gran magnitud** del número de datos (más de 100000 a evaluar), se diseña un **software ad hoc** que nos permita rápidamente **seleccionar las variables** deseadas y organizarlas **una en función de la otra**. Con el lenguaje **Python** y su módulo de **estadística**, se nos facilita el trabajo a la hora de **examinar la base de datos**.

```
class Person:
```

```
    def __init__(self, t, p, ca, s, e, ec, n, pn, ne, al, sp, sxp, hog):
```

Se elabora un análisis descriptivo de los datos mediante **tablas, gráficos, diagramas e histogramas**. Tipificamos en la **normal N(0,1)** y estudiamos los **contrastes de hipótesis**. Con software general como **Microsoft Excel** y específico para estadística como **RStudio**, obtenemos resultados que nos ayudan a sacar conclusiones.

```
        self.sexo, self.edad, self.ecivil, self.pareja = s, e, ec, sxp
```

```
        # Origen del individuo
```

```
        self.nacionalidad, self.pais = n, pn
```

```
        # Estudios y actividades laborales
```

Finalmente, se elaboran y se redactan las **conclusiones** más significativas con respecto al **análisis y estudio** de pasos anteriores, respondiendo a los requisitos planteados en los **objetivos de la investigación**. Se maqueta el trabajo con **PowerPoint e Illustrator** y, para terminar, se revisa y se mejoran los **últimos matices** del proyecto.

```
        self.hogar = hog
```

```
class Hogar:
```

```
    def __init__(self, rt, dc, ch, hab, t, par):
```

1. ANÁLISIS E INTERPRETACIÓN DE DATOS

2. RELACIÓN DE VARIABLES Y OBJETIVOS

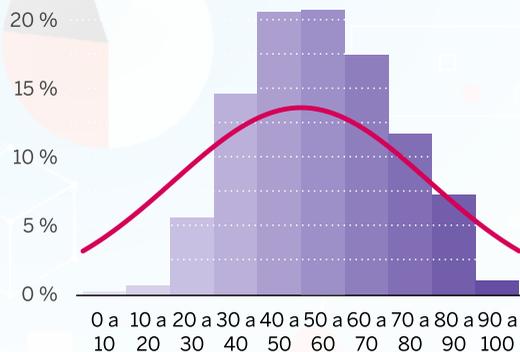
3. COMPUTACIÓN CON CÓDIGO PYTHON PROPIO

4. TABLAS, DIAGRAMAS Y CÁLCULOS

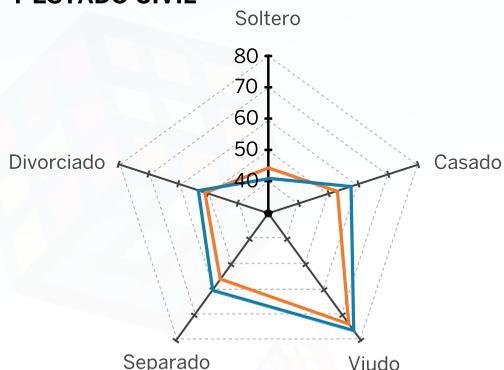
5. CONCLUSIONES, MAQUETACIÓN Y REVISIÓN FINAL

## HISTOGRAMA

Distribución de la variable discreta *Edad*



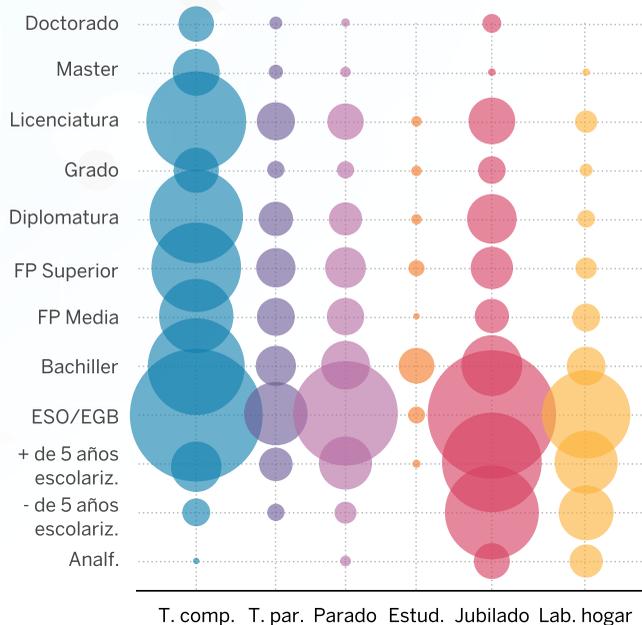
## RELACIÓN ENTRE EDAD MEDIA Y ESTADO CIVIL



## PROTOTIPO DE ENCUESTADO

	Sexo	Edad	Tamaño del municipio	Estado civil	Nacionalidad	Estudios	Situación laboral
Media		55.12					
Valor + frecuente	Mujer		100000-500000	Casada	ESP	ESO / EGB	Tiempo completo
Desv. típica		16.84					
Coef. variación (%)		30.55					

## RELACIÓN ENTRE ESTUDIOS Y ACTIVIDAD LABORAL



## MERCADO LABORAL

24.9% ♂

20.2% ♀

24.5% ♂

30.4% ♀

45.1% de la población es **ACTIVA**

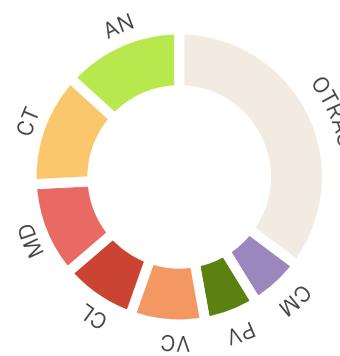
54.9% de la población es **INACTIVA**

# RESULTADOS



VARIABLES DE LA PERSONA ENCUESTADA  
Tablas, diagramas y gráficos

## FRECUENCIA DE ENCUESTADOS POR COMUNIDAD AUTÓNOMA



\* la nomenclatura sigue la norma ISO-3166ES

## SEXO DE LOS ENCUESTADOS



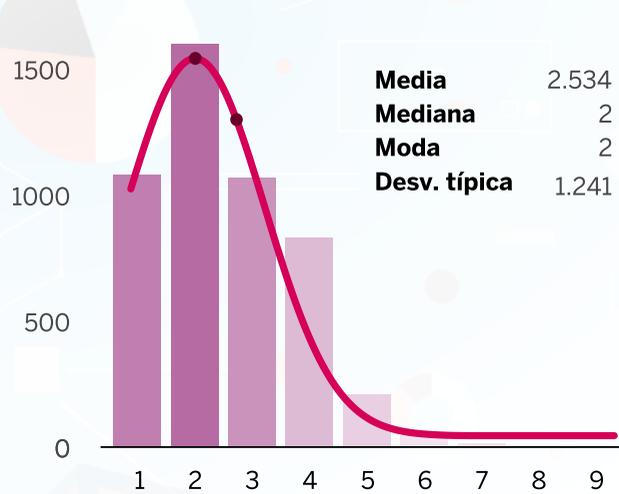
## INFERENCIA ESTADÍSTICA

Con nivel de significación de un 1%, por lo que  $Z_{\alpha/2} = 2.575$

	Contraste de hipótesis	Región de aceptación	Decisión
Ciudad de 100000-500000 hab	$p = 0.2746$	$R(0.2743, 0.2748)$	Se acepta
Estado civil: Viudo	$p = 0.1270$	$R(0.1253, 0.1287)$	Se acepta
Nacionalidad: ESP	$p = 0.9316$	$R(0.9302, 0.9329)$	Se acepta
Estudios: EGB/ESO	$p = 0.3220$	$R(0.3217, 0.3222)$	Se acepta
C.A.: Castilla y León	$p = 0.9316$	$R(0.9302, 0.9329)$	Se acepta

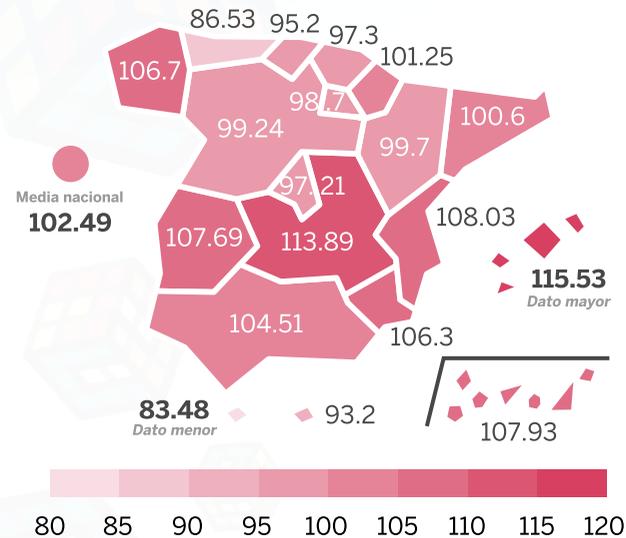
## HABITANTES POR HOGAR

Distribución de frecuencia de dicha variable discreta



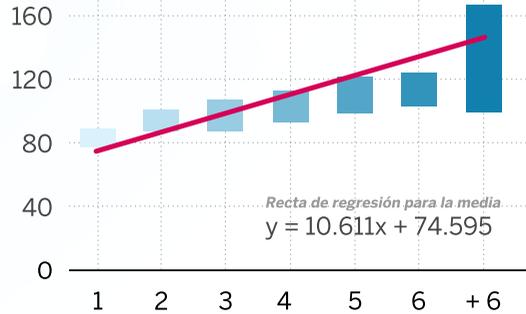
## COMUNIDADES AUTÓNOMAS

Superficie de hogar media en cada C.A. (m<sup>2</sup>)

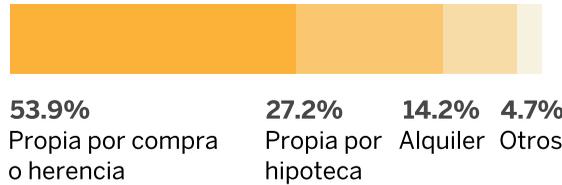


## RELACIÓN ENTRE HABITANTES Y SUPERFICIE DEL HOGAR

El valor inferior corresponde a la mediana, y el superior a la media aritmética



## RÉGIMEN DE TENENCIA DE LA VIVIENDA



## MODELO DE HOGAR

Atendiendo a los parámetros estadísticos habituales

	Régimen de tenencia	Habts.	Superf.	Residentes	Tipo de núcleo familiar	Hijos
Media		5.49	102.49	2.534		0.48
Valor + frecuente	<b>Propiedad</b>				<b>Pareja casada</b>	
Desv. típica		1.49	57.053	1.241		1.21
Coef. variación (%)		27.14	55.66	49.487		252.1

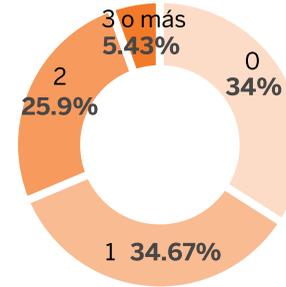
# RESULTADOS



## VARIABLES DEL HOGAR

Tablas, diagramas y gráficos

### NÚMERO DE HIJOS

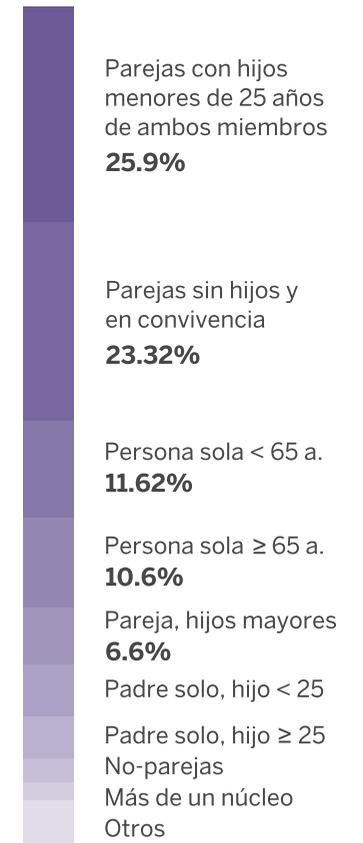


### RÁNKING DE HOGARES POR m<sup>2</sup>

Superficie	Resid.	Provincia
1500 m <sup>2</sup>	3	Guipúzcoa
12 m <sup>2</sup>	1	Gerona

### TIPOS DE HOGAR

Atendiendo núcleos familiares



## RÉGIMEN DE TENENCIA DE LA VIVIENDA

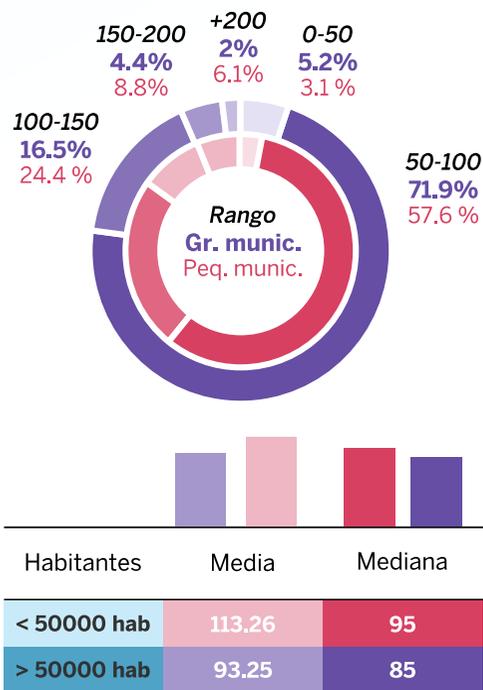
Comparación entre pequeños municipios y grandes municipios



## RELACIÓN ENTRE EDAD Y Nº DE HABITANTES DEL HOGAR



## RELACIÓN ENTRE TAMAÑO DEL MUNICIPIO Y SUPERFICIE DEL HOGAR



# RESULTADOS

CORRELACIÓN entre el encuestado y su hogar

# CONCLUSIONES



## I. CONCLUSIONES SOBRE VARIABLES DEL ENCUESTADO

- En primer lugar, partimos del **histograma** de la variable discreta *Edad*. Teniendo en cuenta que la edad media en España es de 42,92 años según el INE (2017), podemos decir que se produce una **desviación** en este aspecto en la Encuesta de Hogares, cuya media es de 55,12.
- Así mismo, destacamos el **alto grado de variabilidad** de este dato, con una alta desviación típica y un coeficiente de variación de más de 30%.
- Como era de esperar, los solteros/as son, por lo general, más jóvenes; así como los viudos/as son de edad más avanzada. Señalamos también que, salvo en la soltería, la **edad media de la mujer** es **inferior** a la del hombre.
- La diferencia entre la **población activa** y la **inactiva** es muy reducida. Entre la inactiva, predomina la mujer, debido a su tradicional vinculación a las labores domésticas (que no se consideran parte de la economía activa).
- Las **tablas de inferencia** nos permiten comprobar que, a pesar de la desviación con respecto de los valores reales, la Encuesta es **fiable**.
- Según el **gráfico de dispersión**, la población trabajadora en España está hoy más preparada que nunca: observamos que los trabajadores tienen **más nivel de estudio** que sus antecesores, ahora jubilados.
- La mayoría de los **trabajadores** tienen la **EGB**, implementada por la Ley de Educación de **1970**; mientras que en los **jubilados** es más frecuente la **escolarización de 5 a 10 años** en base a la Ley Moyano de 1857.
- Señalamos que el **top 3** de la Encuesta por autonomías coincide con las tres CCAA más pobladas, i.e. **Andalucía, Cataluña y Madrid**.

# CONCLUSIONES



## II. CONCLUSIONES SOBRE VARIABLES DEL HOGAR

- La gráfica de **habitantes por hogar** nos permite observar que muy pocos hogares tienen más de cuatro residentes, de ahí la **cercanía entre la media, la mediana y la moda**. La desv. típica es muy reducida (1,241)
- Podemos destacar la **división Norte-Sur** de España en cuanto a la superficie de hogar media, albergando el sur una media de superficie más alta que el norte. Esto se debe a **dos razones**: una **sociohistórica**, ya que tradicionalmente y más desde el siglo XIX el **régimen de propiedad** en el sur de España ha sido menos limitado que en el norte (hay más ensanches septentrionales que meridionales); y por otro lado, por una razón **orográfica**: al ser el sur de España menos escarpado que el norte, se motiva la **construcción horizontal** con más m<sup>2</sup> útiles.
- Así mismo, la relación entre **habitantes y superficie del hogar** es lineal (sigue aproximadamente la **recta de regresión** indicada). Sin embargo, cuando el número de residentes excede de los 6, se produce una **variación mayor** entre la media (afectada por los valores extremos) y la mediana, tal y como se puede ver en el gráfico.
- A diferencia de otros países de la Unión Europea, el **régimen de tenencia** predominante es la propiedad privada, ya sea por compra o por hipoteca.
- Los datos de hijos por hogar nos indican la pertenencia de España a un **régimen demográfico moderno** de Fase 4 o Fase 5 (crecimiento cero)
- En cuanto a la clasificación por **tipos de hogar**, destacamos que la gran mayoría tienen una **relación de parentesco** entre sus miembros. Sin embargo, más allá de lo numérico, me choca y me preocupa la **alta tasa** de personas **mayores de 65 años** que viven completamente **solos** (10,6%). Me pregunto si, realmente, esa situación se puede llamar hogar.

## III. CONCLUSIONES SOBRE CORRELACIONES ENTRE VARIABLES

- Considerando **pequeño municipio** como aquel de menos de 5000 habitantes, podemos afirmar que existe una **alta concentración urbana** entre los encuestados. La única diferencia notable en esta división es el **aumento de los alquileres urbanos**, motivado por la crisis de 2008.
- En los grandes municipios existe un **predominio de viviendas** con una superficie entre los 50 y los 100 m<sup>2</sup>. Se debe, una vez más, a la urbanística del s. XIX: por ejemplo, las viviendas del Eixample y los barrios obreros en Barcelona tienen superficies en torno a los 50 m<sup>2</sup> (Univ. de Barcelona). Destaca la concentración de **grandes viviendas en municipios pequeños**.
- Por último, señalamos la **escasa relación** entre la edad y el número de habitantes de un hogar, ya que, en todos los casos, tanto la **media** como la **mediana** se concentran entre el 5 y el 5,6 pers./hogar.

## IV. CONCLUSIONES FINALES DEL PROYECTO

- Aunque la gran cantidad de entradas en la Encuesta (n=5000) invita al desasosiego, no se han encontrado dificultades a la hora de **procesar los datos**. Se echan en falta variables como los **ingresos medios por hogar** o la **antigüedad** de la vivienda, que se hubieran podido relacionar con otros datos para obtener más conclusiones y reforzar las existentes.
- La elaboración de **gráficas y diagramas**, así como la inferencia, han sido de gran ayuda para sacar conclusiones interesantes en temas variados.
- En vista de todo lo expuesto, este proyecto ha servido para ampliar nuestra **altura de miras** con respecto del **tema propuesto**, así como nos ha ofrecido una **nueva perspectiva** y nuevos conocimientos interesantes sobre el apasionante campo de la **estadística** y la **demografía**.