

# La Encuesta de Condiciones de Vida y la vivienda

Javier Martínez Santos

## Introducción

Lo primero que hacemos es cargar los paquetes necesarios

```
library(tidyverse)
library(dineq)
library(ariamsita)
library(viridis)
library(viridisLite)
library(stargazer)
library(sjPlot)
```

En este artículo se han obtenido una serie de resultados a partir de la Encuesta de Condiciones de Vida que publica el INE periódicamente de forma anual. Se han unificado los ficheros P y R con información personal a partir de las variables:  $RB030 = PB030$ , luego los ficheros H y D referentes a los hogares a partir de las variables:  $DB030 = HB030$  y por último se han agregado ambos dataframes usando las dos variables resultantes de los ficheros y renombrada a `id_hh`. Este proceso se ha llevado a cabo para las encuestas de 2014, 2017 y 2020.

Así hemos obtenido un fichero único con información de las personas, es decir, de manera individual, pero incorporando también la información de las mismas personas en referencia a las preguntas sobre el hogar.

A continuación se expondrán los resultados, primero un análisis univariante, luego un análisis bivariante con información espacial y por último un modelo logit binomial multivariante.

```
load("ecv.RData")
```

## Régimen de tenencia

Lo primero que se ha hecho es seleccionar las variables que se iban a usar, en especial: el régimen de tenencia, el año de la encuesta, los pesos o Factor de ponderación transversal de la persona y el año de nacimiento para calcular la edad de la persona.

Luego se han codificado las respuestas de la variable “régimen” de factores numéricos a factores caracteres.

Por último se ha calculado el porcentaje de personas que viven en cada régimen, agrupando primero por las variables “régimen” y “año” para calcular  $\text{sum}(\text{pesos})$  para saber cuantas personas hay en cada régimen para cada año concreto. Posteriormente, agrupando por cada año se calcula el porcentaje de personas por régimen de tenencia y año.

```
df1 <- ecv %>% select(id_hh, ccaa=DB040, regimen=HH021, año=DB010, pesos = RB050)

df1 <- df1 %>% mutate(regimen = case_when(
  regimen == 1 ~ "En propiedad sin hipoteca",
  regimen == 2 ~ "En propiedad con hipoteca",
  regimen == 3 ~ "En alquiler de mercado",
  regimen == 4 ~ "En alquiler inferior al de mercado",
  regimen == 5 ~ "En cesión gratuita"))

df1 <- df1 %>%
```

```

group_by(regimen, año) %>%
summarise(n = sum(pesos)) %>%
group_by(año) %>%
mutate(pct = n/sum(n)*100) %>%
mutate(pct = round(pct, 2))

df1$año <- as.factor(df1$año)

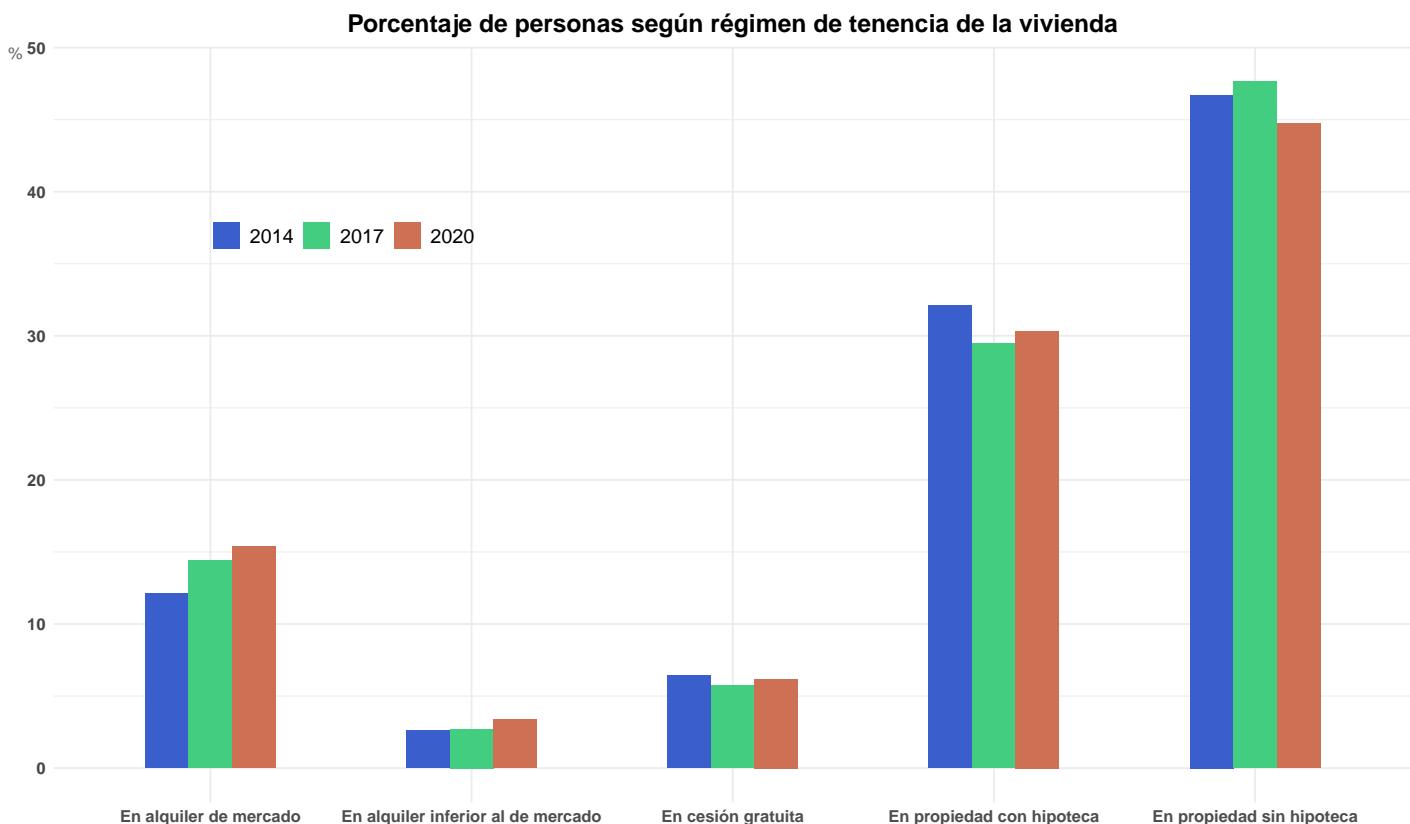
```

Gracias a estos cálculos y la representación gráfica posterior, se puede observar que en torno a un 47 de la población española reside en viviendas en propiedad sin hipoteca, este porcentaje esta influido en gran parte por los mayores de 60 años, grupo en el cual mas del 80% reside en este régimen de tenencia y estos son aproximadamente 11,5 millones de personas en 2019.

```

df1 %>% ggplot(aes(x=regimen, y = pct, fill = año)) +
  geom_bar(stat = "identity", position=position_dodge() , width = 0.5) +
  labs(title = "Porcentaje de personas según régimen de tenencia de la vivienda",
       x = "",
       y = "%") +
  scale_fill_manual(values = c("royalblue3", "seagreen3", "salmon3")) +
  theme_ariamsita() +
  theme(axis.title.x = element_blank(),
        axis.text.x = element_text(face = "bold", size = 10),
        axis.text.y = element_text(face = "bold", size = 10),
        plot.title = element_text(size = 15),
        legend.title = element_blank(),
        legend.position = c(0.21, 0.75),
        legend.direction = "horizontal",
        legend.text = element_text(size = 12))

```



## Alquiler por Comunidades Autónomas

A continuación, como ejemplo de análisis bivariante se ha calculado de nuevo el porcentaje de personas que viven en régimen de alquiler en España, pero esta vez por comunidades autónomas. Para la representación gráfica nos hemos ayudado del paquete MapSpain, el cual contiene las distintas geometrías de España, tanto a nivel estatal, autonómico, provincial y municipal.

Posteriormente, se han calculado los porcentajes de personas que viven en régimen de alquiler en el año 2020 agrupando por la CCAA en la que residen dichas personas.

```
df2 <- ecv %>% select(id_hh, ccaa=DB040, regimen=HH021, año=DB010, pesos = RB050) %>%
  filter(año == 2020)
df2 <- df2 %>% mutate(regimen = ifelse(regimen %in% 3:4, "En alquiler", "En propiedad"))

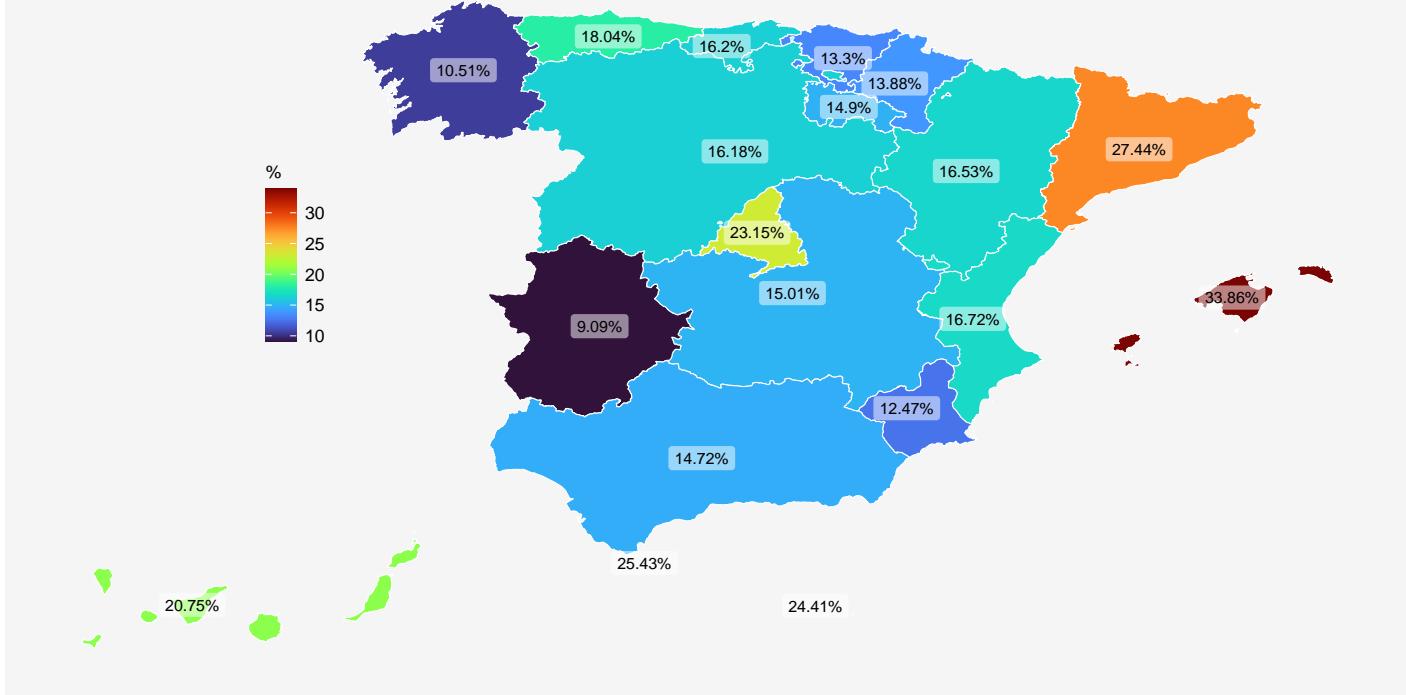
df2 <- df2 %>%
  group_by(regimen, ccaa) %>%
  summarise(n = sum(pesos)) %>%
  group_by(ccaa) %>%
  mutate(pct = n/sum(n)*100) %>%
  mutate(pct = round(pct, 2))

df2$pct.lab <- paste0(df2$pct, "%")

plot <- mapSpain::esp_get_ccaa()
df2 <- df2 %>% left_join(plot, by = c("ccaa" = "nuts2.code"))
df2 <- df2 %>% filter(regimen == "En alquiler")

df2 %>% ggplot() +
  geom_sf(aes(fill = pct, geometry = geometry), color = "white", lwd = 0.3) +
  scale_fill_viridis(option = "turbo", name = "%") +
  geom_sf_label(aes(label = pct.lab, geometry = geometry),
                fill = "white", alpha = 0.5, size = 3, label.size = 0) +
  labs(title = "Porcentaje de personas que viven en alquiler por CCAA en 2020") +
  theme_minimal() +
  theme(
    axis.line = element_blank(),
    axis.text.x = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks = element_blank(),
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    plot.background = element_rect(fill = "grey96", color = NA),
    panel.background = element_rect(fill = "grey96", color = NA),
    legend.background = element_rect(fill = "grey96", color = NA),
    panel.border = element_blank(),
    plot.title = element_text(size = 15, hjust = 0.5, face = "bold"),
    legend.title = element_text(color = "black", size = 10),
    legend.text = element_text(color = "black", size = 10, hjust = 0),
    legend.position = c(0.2, 0.6))
```

## Porcentaje de personas que viven en alquiler por CCAA en 2020



Como se ve en el mapa, hay gran disparidad entre las Comunidades Autónomas en relación al porcentaje de personas que viven en alquiler en dichos territorios. Por ejemplo, nos encontramos el caso de Galicia, donde únicamente viven en alquiler el 10,5% de sus habitantes o Extremadura donde este valor cae hasta el 9%, mientras que Cataluña o Baleares alcanzan el 27,4% y 33,8% respectivamente.

## Modelo logístico

A continuación, se ha estimado un modelo logístico con el objetivo de estimar la probabilidad de residir en una vivienda en régimen de propiedad (con y sin hipoteca). Para la creación de este modelo de elección se ha seleccionado la variable “Renta” calculada como  $\text{elog}(\text{vhRentaa}/\text{HX240})$ , es decir, el logaritmo del ingreso equivalente por unidad de consumo, luego se ha codificado la variable “cohorte” en 4 grupos según la década de nacimiento de la persona, desde 1955 hasta 1994 y como última variable se ha añadido el percentil de renta calculado en base al ingreso equivalente, en euros, mencionado anteriormente.

```
df3 <- ecv %>%
  select(año = DB010, renta = vhRentaa, HX240, tamaño.hogar = HX040,
         cohorte = RB080, sexo = RB090, situacion = RB210,
         estudios = PE040, regimen = HH021, urbanizacion = DB100, RB050) %>%
  mutate(renta = renta/HX240,
         Percentil = ntile.wtd(renta, n = 100, RB050),
         renta = log(renta)) %>%
  filter(cohorte %in% 1955:1994, renta >0, año == 2020) %>%
  mutate(cohorte = case_when(
    cohorte %in% 1955:1964 ~ "1955-1964",
    cohorte %in% 1965:1974 ~ "1965-1974",
    cohorte %in% 1975:1984 ~ "1975-1984",
    cohorte %in% 1985:1994 ~ "1985-1994"))

df3 <- df3 %>% mutate(regimen = ifelse(regimen %in% 1:2, "1", "0"),
                         estudios = ifelse(estudios>=500, 1, 0))
```

```

df3$regimen <- as.factor(df3$regimen)
df3$cohorte <- as.factor(df3$cohorte)

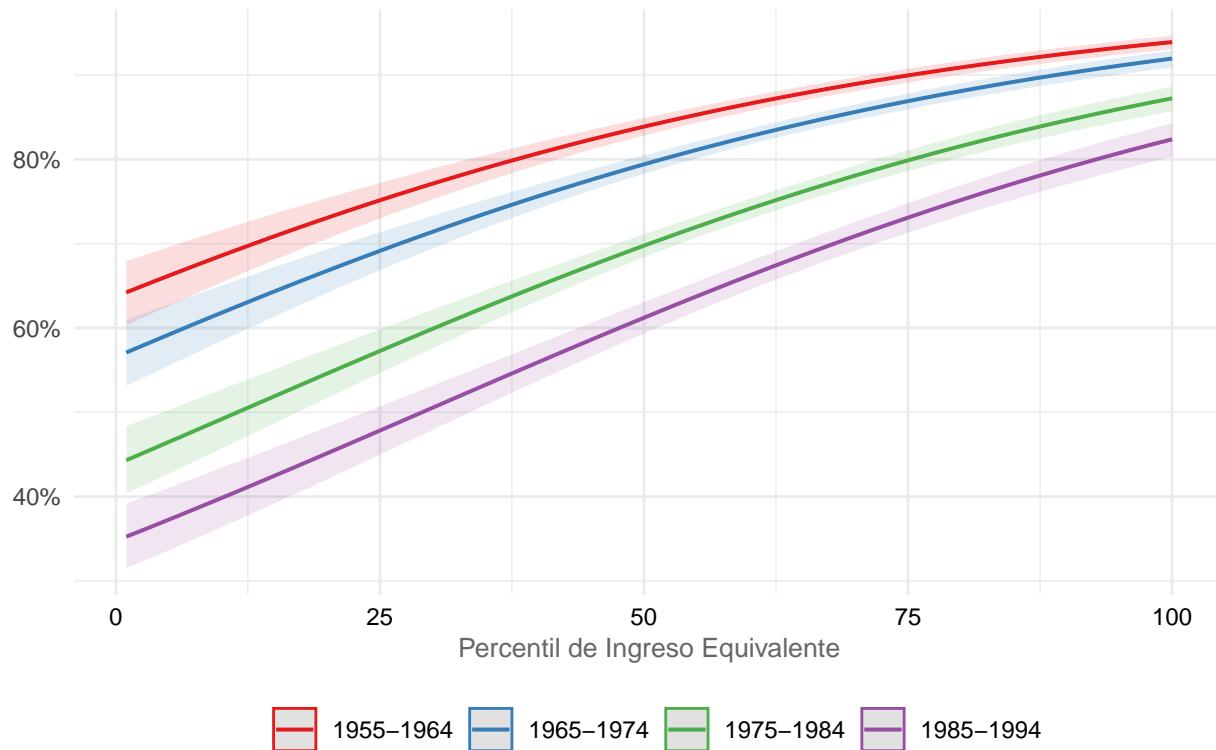
model <- glm(family=binomial(link="logit"), data = df3,
             regimen ~ renta + cohorte + Percentil)

plot_model(model,type = "pred",
           terms=c("Percentil[all]", "cohorte[all]")) + geom_line(size = .5) +
  labs(title = "Probabilidad de tener una vivienda en propiedad por percentil de renta en 2020",
       subtitle = "Por cohorte de edad",
       x = "Percentil de Ingreso Equivalente",
       y = "")+
  theme_ariamsita()+
  theme(plot.title = element_text(size = 10),
        axis.title.x = element_text(size = 10),
        axis.text = element_text(size = 10, color = "black"),
        legend.position = "bottom",
        legend.title = element_blank())

```

### Probabilidad de tener una vivienda en propiedad por percentil de renta en 2020

Por cohorte de edad



Como se ve en el gráfico, la probabilidad de residir en una vivienda en propiedad va disminuyendo a medida que aumenta la década de nacimiento de las personas, perteneciendo al mismo percentil de la renta.

Las personas nacidas en la década de 1985 y que pertenecen al percentil 50 de la distribución, tiene aproximadamente un 20% menos de probabilidad de residir en una vivienda en propiedad frente a los nacidos en la década de 1955.

Esto puede indicarnos, a priori, un problema de acceso a la vivienda para las generaciones más jóvenes, eliminando el factor renta, ya que como hemos visto, a mismo percentil la probabilidad de residir en una vivienda en propiedad sigue siendo menor.

```
stargazer(model, title = "Logit estimation, Log Odds", header = FALSE)
```

Table 1: Logit estimation, Log Odds

<i>Dependent variable:</i>	
	regimen
renta	−0.006 (0.052)
cohorte1965-1974	−0.300*** (0.050)
cohorte1975-1984	−0.813*** (0.049)
cohorte1985-1994	−1.193*** (0.054)
Percentil	0.022*** (0.001)
Constant	0.624 (0.434)
Observations	20,756
Log Likelihood	−10,233.550
Akaike Inf. Crit.	20,479.100

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```
stargazer2(model, odd.ratio = T,title = "Logit estimation, Odds ratios", header = FALSE)
```

Por ultimo, se encuentran las Tablas 1 y 2 que muestran los resultados del modelo logístico explicado anteriormente. La primera tabla hace referencia a los resultados en Log Odds, lo que hace complicado poder interpretarla de manera sencilla, por lo que, en la segunda tabla aparecen calculados los Odds Ratios.

Table 2: Logit estimation, Odds ratios

<i>Dependent variable:</i>	
	regimen
renta	0.994 (0.051)
cohorte1965-1974	0.741*** (0.037)
cohorte1975-1984	0.443*** (0.022)
cohorte1985-1994	0.303*** (0.016)
Percentil	1.022*** (0.001)
Constant	1.866 (0.810)
Observations	20,756
Log Likelihood	-10,233.550
Akaike Inf. Crit.	20,479.100

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01