

Análisis del 10% de mayores ingresos en Chile

Marcos González, Agustín Rabie

19 January, 2025

Contents

Este proyecto analiza la distribución y características del 10% superior de los hogares de mayores ingresos en Chile, utilizando datos de las encuestas CASEN 2020 y EBS 2021. El objetivo es desarrollar un análisis descriptivo detallado y posteriormente un modelo de machine learning que permita predecir la pertenencia a este grupo.

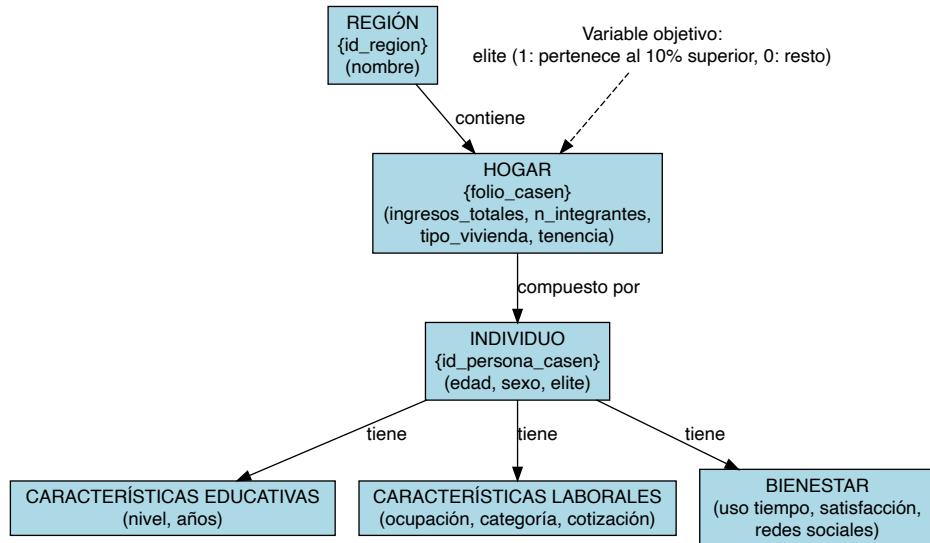
1. Preparación y descripción de los datos Para comenzar el análisis, se cargan las bibliotecas necesarias y se establecen los parámetros de configuración. Se utiliza una combinación de paquetes para manipulación de datos (dplyr), visualización (ggplot2), manejo de datos espaciales (sf, geodata, chilemapas) y presentación de resultados (kableExtra).

2. Carga y preparación inicial de datos El análisis utiliza dos fuentes principales de datos:

CASEN 2020 (versión reducida, en formato rds para poder subirse a GitHub): Proporciona información socioeconómica detallada EBS 2021: Complementa con información adicional y factores de expansión actualizados

En el proceso de preparación, se realizan los siguientes pasos: a. Cálculo de deciles de ingreso para identificar el 10% superior según la muestra total de la encuesta CASEN, no de la submuestra de la EBS b. Creación de variable binaria para el grupo objetivo c. Merge de ambas bases de datos manteniendo la estructura de la CASEN

3. Modelo Entidad-Relación La estructura de datos se organiza en cuatro entidades principales:



4. Análisis descriptivo inicial Para comprender la distribución de ingresos en Chile, comenzamos examinando las estadísticas básicas por decil. Esto nos permite contextualizar el 10% superior dentro de la

Table 1: Estadísticas por decil de ingreso

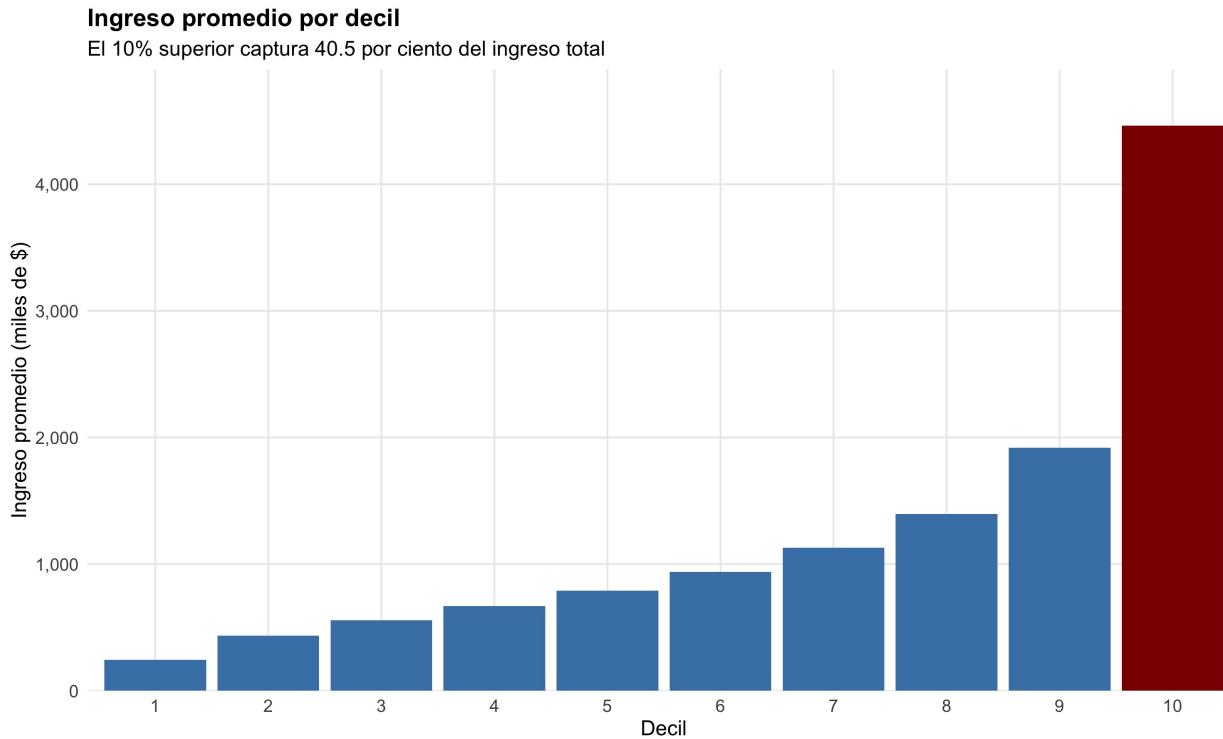
decil	n	n_expandido	ingreso_medio	proporcion_poblacion
1	1117	1325185	243723.2	0.0873880
2	1204	1499566	436750.2	0.0988873
3	1099	1389221	558032.0	0.0916107
4	1093	1443258	667421.0	0.0951742
5	1096	1423513	790322.5	0.0938721
6	1082	1541201	937046.2	0.1016329
7	1100	1597200	1130165.5	0.1053257
8	1104	1533102	1397455.4	0.1010988
9	1052	1552609	1917867.4	0.1023852
10	974	1859535	4460647.5	0.1226251

distribución general de ingresos del país.

Tenemos, para nuestro caso positivo, 974 casos de una muestra de 10921, sin NAs, esto es, alrededor de un 9% de la muestra, lo que se considera suficiente para informar un modelo de Machine Learning.

5. Visualización de la distribución por deciles El gráfico siguiente muestra la marcada desigualdad en la distribución de ingresos en Chile. Se destaca el porcentaje del ingreso total que captura ese 10%.

```
## pdf
## 2
```



6. Caracterización del 10% superior vs resto de la población La tabla siguiente compara características clave entre este grupo y el resto de la población. Se incluyen variables territoriales (región, urbano/rural), demográficas (edad, sexo), socioeconómicas (educación) y de bienestar. Los valores están

ponderados usando los factores de expansión provistos por la EBS: seleccionamos éstos en vez de los de CASEN por tratarse de una encuesta bifásica.

\begin{table} \caption{Características del 10% superior vs resto}

elite_label	n	n_expandido	ingreso_promedio	desv_est	edad_promedio	prop_hombres	prop_urbano
10% superior	974	1,859,535	4,460,648	2,993,049	41.27	0.56	0.95
Resto	9,947	13,304,855	917,774	500,956	44.87	0.48	0.87

\end{table}

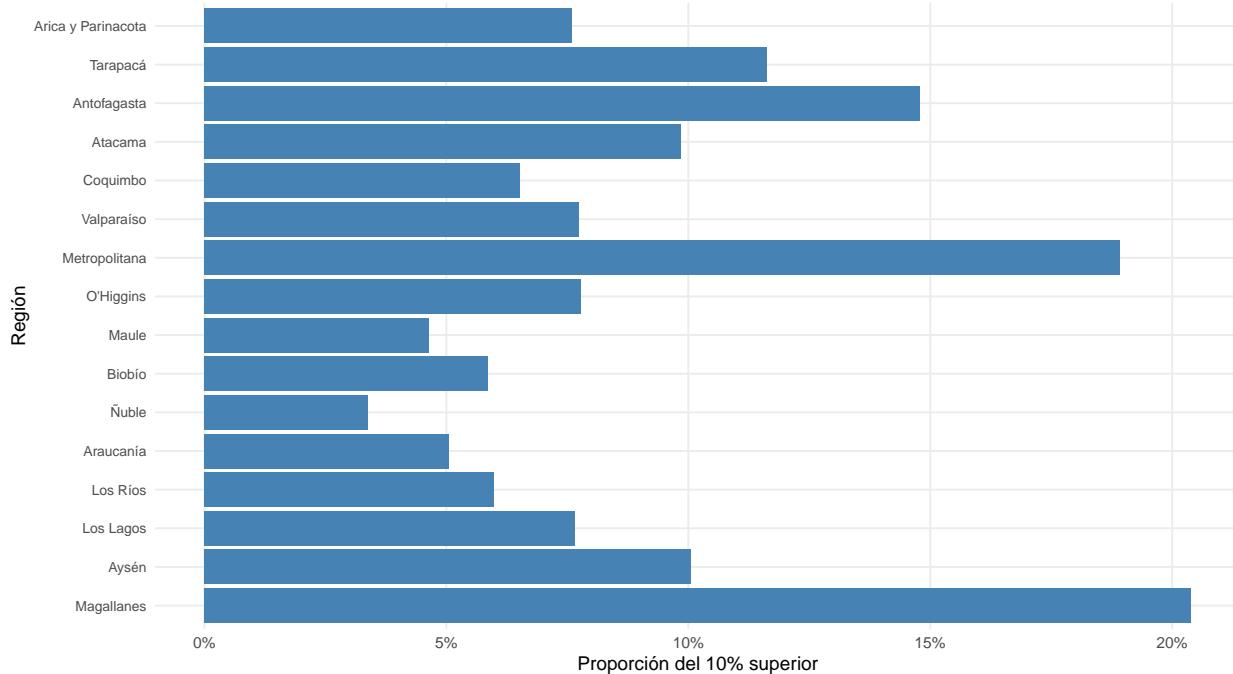
7. Análisis territorial El análisis territorial se desarrolla en dos dimensiones complementarias:

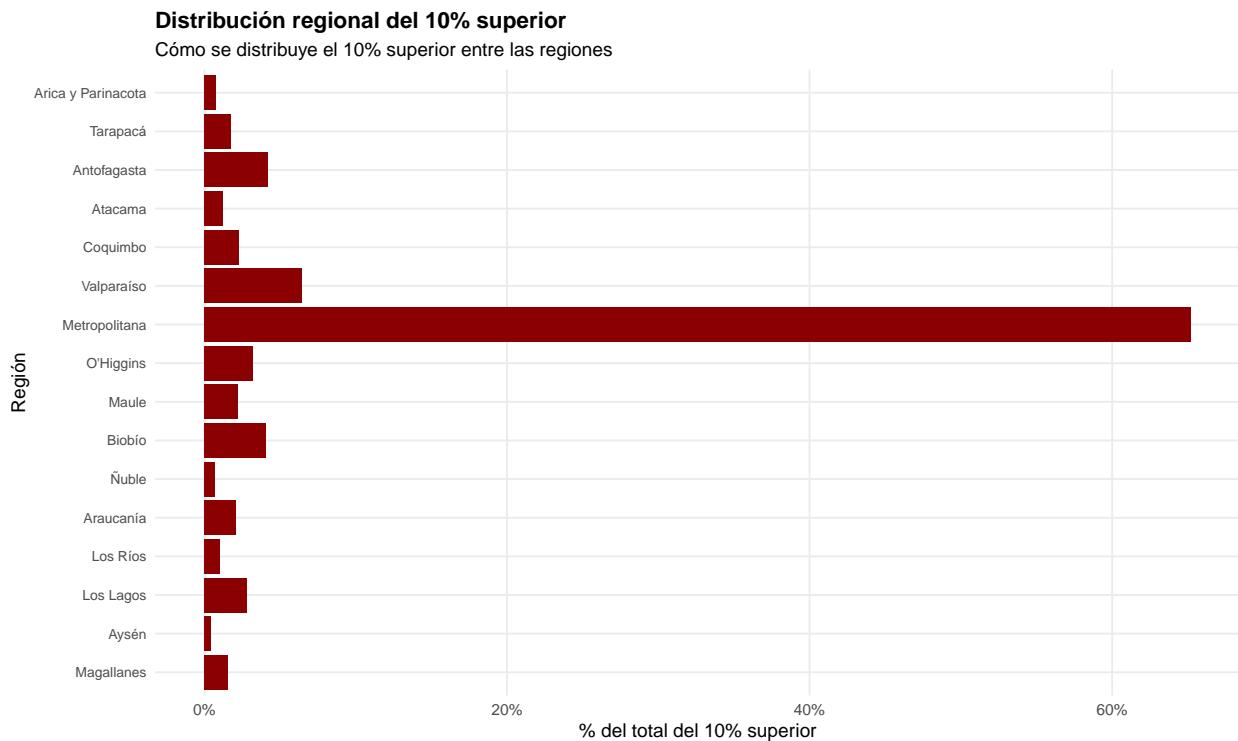
Proporción interna: Qué porcentaje de la población de cada región pertenece al 10% superior Distribución nacional: Cómo se distribuye el total del 10% superior entre las regiones

Para facilitar la interpretación, se ordenan las regiones de norte a sur.

Proporción del 10% superior por región

% de habitantes de cada región que pertenece al 10% superior





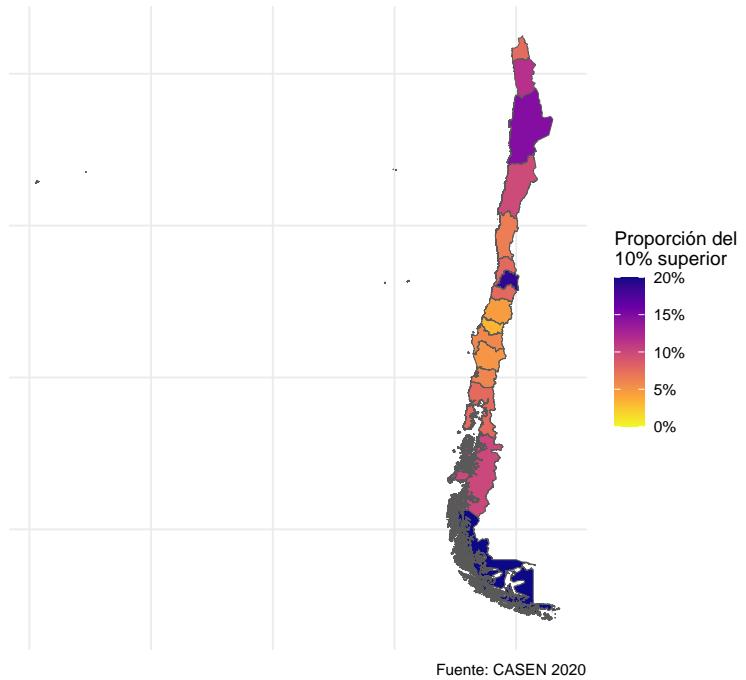
Los gráficos resultantes revelan patrones interesantes. La proporción de elite dentro de cada región (gráfico azul) muestra una concentración en ciertas áreas. La distribución del total de la elite (gráfico rojo) evidencia una fuerte centralización

8. Visualización espacial Aquí generamos mapas que permiten una visualización más intuitiva de los patrones espaciales. Se utilizan dos mapas que corresponden a las mismas dimensiones analizadas en los gráficos de barras:

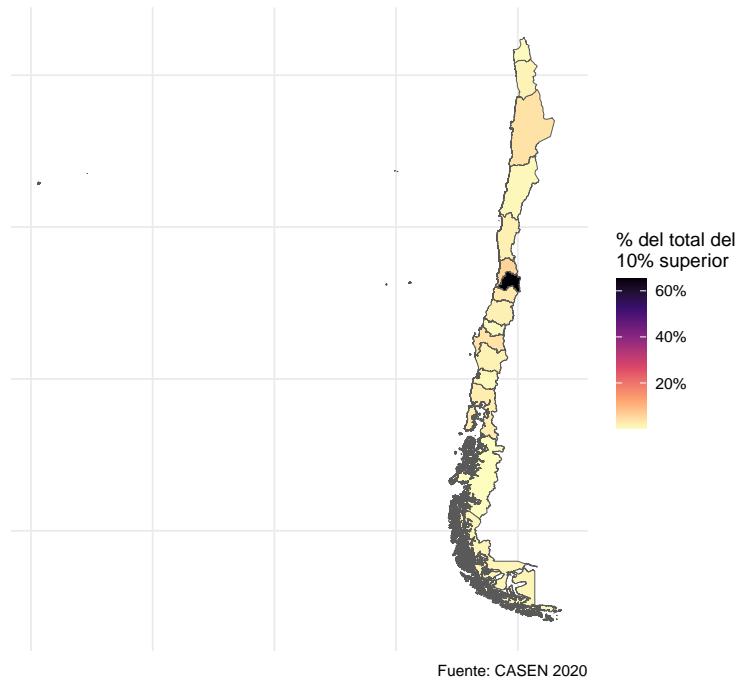
El primer mapa muestra la proporción de habitantes de cada región que pertenece al 10% superior. Este mapa ayuda a identificar dónde es más probable encontrar miembros de la elite económica. El segundo mapa (en tonos magma) visualiza cómo se distribuye el total de ese 10% entre las regiones.

Para la construcción de estos mapas, se enfrentaron varios desafíos técnicos: La necesidad de compatibilizar diferentes codificaciones de regiones: Magallanes tenía problemas de visualización que lo convertía persistentemente en NA.

Proporción del 10% superior por región
% de habitantes de cada región que pertenece al 10% superior



Distribución regional del 10% superior
Cómo se distribuye el 10% superior entre las regiones



9. Sexo jefe/a de hogar

Recodificación para identificar sexo de jefe de hogar. No es posible saberlo para el 100% de los casos, pero con variables de sexo y de relación de quien responde con jefe/a de hogar, se puede reducir para cerca del 70% de los casos.

Table 2: Distribución de parentesco

Parentesco	Frecuencia
Jefe(a) de Hogar	5164
Esposo(a) o pareja de distinto sexo	2442
Esposo(a) o pareja de igual sexo	18
Hijo(a) de ambos	1078
Hijo(a) sólo del jefe(a)	1331
Hijo(a) sólo del esposo(a)/pareja	65
Padre o madre	123
Suegro(a)	34
Yerno o nuera	119
Nieto(a)	234
Hermano(a)	138
Cuñado(a)	20
Otro Familiar	98
No familiar	57

Table 3: Distribución y proporción de elite por sexo del jefe de hogar

Sexo	N	N expandido	Proporción elite	Proporción muestra	Proporción población
Mujer	2710	3129383	6.2%	0.525	0.454
Hombre	2454	3757075	14.1%	0.475	0.546

```
## [1] "\nDistribución del sexo del jefe de hogar (incluyendo inferidos):"
##
##      1    2 <NA>
## 4357 3267 3297
```

Sexo del/la jefe de hogar

Edad

10. Nivel educacional

Resumen variables potencialmente predictoras Para preparar la fase de modelamiento, se realiza un análisis exploratorio de las variables que podrían predecir la pertenencia al 10% superior. El análisis de estas variables se realiza considerando:

Su distribución diferenciada entre elite y no elite La presencia de valores faltantes que podrían afectar el modelamiento La necesidad de transformaciones o recodificaciones para su uso en modelos predictivos

Variables potenciales

La tabla resultante proporciona una primera aproximación a la capacidad predictiva de cada variable, mostrando diferencias significativas en varias dimensiones entre el grupo elite y el resto de la población.

Variables de bienestar de EBS

Table 4: Distribución y proporción de elite por grupo de edad

Grupo de edad	N	N expandido	Proporción elite (ponderada)	Proporción muestra	Proporción población
18-29	293	435288	2.2%	0.057	0.063
30-34	397	670537	19.5%	0.077	0.097
35-39	433	694170	17.3%	0.084	0.101
40-44	486	770243	13.2%	0.094	0.112
45-49	564	674818	14.3%	0.109	0.098
50-54	584	717096	9.3%	0.113	0.104
55-59	593	782963	8.9%	0.115	0.114
60-64	563	625209	8.7%	0.109	0.091
65-69	466	552611	6.5%	0.090	0.080
70-74	361	436181	5.5%	0.070	0.063
75-79	248	282900	5.4%	0.048	0.041
80 o más	176	244442	0.3%	0.034	0.035

Table 5: Distribución y proporción de elite por nivel educacional

Nivel educacional	N	N expandido	Proporción elite (ponderada)	Proporción muestra	Proporción población
Hasta básica	2453	2879582	1.3%	0.225	0.190
Hasta media	4613	6227468	4.3%	0.422	0.411
Postgrado	188	334451	58.7%	0.017	0.022
Técnica superior	1302	1879090	9.9%	0.119	0.124
Universitaria	2365	3843799	30.5%	0.217	0.253

Table 6: Distribución y proporción de elite por tenencia de vivienda

Tenencia	N	N expandido	Proporción elite (ponderada)	Proporción muestra	Proporción población
Propia pagada	6022	7424366	9.7%	0.551	
Propia pagándose	1052	2100421	36.1%	0.096	
Propia compartida (pagada)	39	65280	12.9%	0.004	
Propia compartida (pagándose)	3	6793	88.4%	0.000	
Arrendada	2097	3307550	8.4%	0.192	
Cedida	1305	1707647	3.9%	0.119	
Usufructo	301	406743	5.1%	0.028	
Ocupación irregular	74	111732	0.2%	0.007	
Poseedor irregular	28	33858	0.0%	0.003	

Table 7: Estadísticas de edad por grupo

elite_label	N	Media	DE	NA's (%)
10% superior	974	41.27	16.05	0%
Resto	9947	44.87	17.84	0%

Table 8: Distribución por sexo

sexo	N sin ponderar	N ponderado	N elite	% del total	% Elite (ponderado)	NA's (%)
Mujer	6308	7753344	477	57.8%	10.6%	0%
Hombre	4613	7411046	497	42.2%	14.0%	0%

Table 9: Distribución por zona

zona	N sin ponderar	N ponderado	N elite	% del total	% Elite (ponderado)	NA's (%)
Urbana	9307	13405487	916	85.2%	13.2%	0%
Rural	1614	1758903	58	14.8%	5.3%	0%

Table 10: Distribución por región

region_nombre	N sin ponderar	N ponderado	N elite	% del total	% Elite (ponderado)	NA's (%)
Metropolitana	1138	6412096	205	10.4%	18.9%	0%
Valparaíso	851	1541851	66	7.8%	7.7%	0%
Biobío	761	1289286	39	7.0%	5.9%	0%
Araucanía	664	776319	31	6.1%	5.1%	0%
Ñuble	645	399028	23	5.9%	3.4%	0%
O'Higgins	644	766289	40	5.9%	7.8%	0%
Los Lagos	633	687519	48	5.8%	7.7%	0%
Los Ríos	631	315343	34	5.8%	6.0%	0%
Coquimbo	623	641782	33	5.7%	6.5%	0%
Antofagasta	623	527209	85	5.7%	14.8%	0%
Tarapacá	623	285450	68	5.7%	11.6%	0%
Atacama	621	233615	57	5.7%	9.9%	0%
Maule	620	876150	26	5.7%	4.6%	0%
Aysén	618	80017	67	5.7%	10.1%	0%
Arica y Parinacota	617	191704	47	5.6%	7.6%	0%
Magallanes	609	140732	105	5.6%	20.4%	0%

Table 11: Distribución por nivel educacional

educ_rec	N sin ponderar	N ponderado	N elite	% del total	% Elite (ponderado)	NA's (%)
Hasta media	4613	6227468	184	42.2%	4.3%	0%
Hasta básica	2453	2879582	34	22.5%	1.3%	0%
Universitaria	2365	3843799	535	21.7%	30.5%	0%
Técnica superior	1302	1879090	125	11.9%	9.9%	0%
Postgrado	188	334451	96	1.7%	58.7%	0%

Table 12: Distribución por tenencia de vivienda

tenencia_vivienda	N sin ponderar	N ponderado	N elite	% del total	% Elite (ponderado)	NA's (%)
Propia pagada	6022	7424366	446	55.1%	9.7%	0%
Arrendada	2097	3307550	131	19.2%	8.4%	0%
Cedida	1305	1707647	51	11.9%	3.9%	0%
Propria pagándose	1052	2100421	329	9.6%	36.1%	0%
Usufructo	301	406743	13	2.8%	5.1%	0%
Ocupación irregular	74	111732	1	0.7%	0.2%	0%
Propria compartida (pagada)	39	65280	2	0.4%	12.9%	0%
Poseedor irregular	28	33858	0	0.3%	0.0%	0%
Propria compartida (pagándose)	3	6793	1	0.0%	88.4%	0%

Table 13: Comparación de variables de bienestar entre grupos (valores estandarizados)

Variable	Media no elite	DE no elite	Media elite	DE elite
Apoyo a empleabilidad	-0.03	1.00	0.20	0.95
Balance trabajo-vida	-0.02	1.00	0.14	0.96
Satisfacción con tiempo	-0.01	1.00	0.05	1.02
Logro de metas	-0.05	1.01	0.40	0.84
Interferencia doméstica	0.01	1.01	-0.09	0.95
Flexibilidad ausencias	0.78	0.42	0.83	0.38