

IX Congreso Andaluz de Sociología  
Huelva 22 de Noviembre

## Taller 2: Paneles online. Tipos y aplicaciones. El Panel Ciudadano para la Investigación Social en Andalucía (PACIS)

Juan Antonio Domínguez  
Manuel Trujillo



# **Paneles probabilísticos on-line. Reclutamiento, calidad y mantenimiento**

1. Paneles online. Conceptos básicos
2. Reclutamiento de hogares para panel probabilístico
3. Evaluación de la representatividad del panel
4. Realización de las encuestas
5. Mantenimiento del panel

# Paneles online. Conceptos básicos

1. Definición
2. Historia
3. Tipos de paneles online
4. Calidad de los datos

### PANEL DE ENCUESTAS EN GENERAL

**Muestras pre-reclutadas de la población para que participe en encuestas periódicas.**

### PANEL ONLINE EN PARTICULAR

**“Bases de datos de encuestados potenciales que se declaran dispuestos a cooperar en encuestas futuras si son seleccionados para ello”** (Callegaro, Lozar Manfreda y Vehovar)

Los tipos de paneles varían según el universo de informantes y/o el tipo de información recogida. Destacando el uso para medición de **audiencias** y **de consumo**. Hay que distinguir además entre paneles **probabilísticos** y **no probabilísticos**

# Paneles online

## HITOS EN CREACIÓN DE PANELES

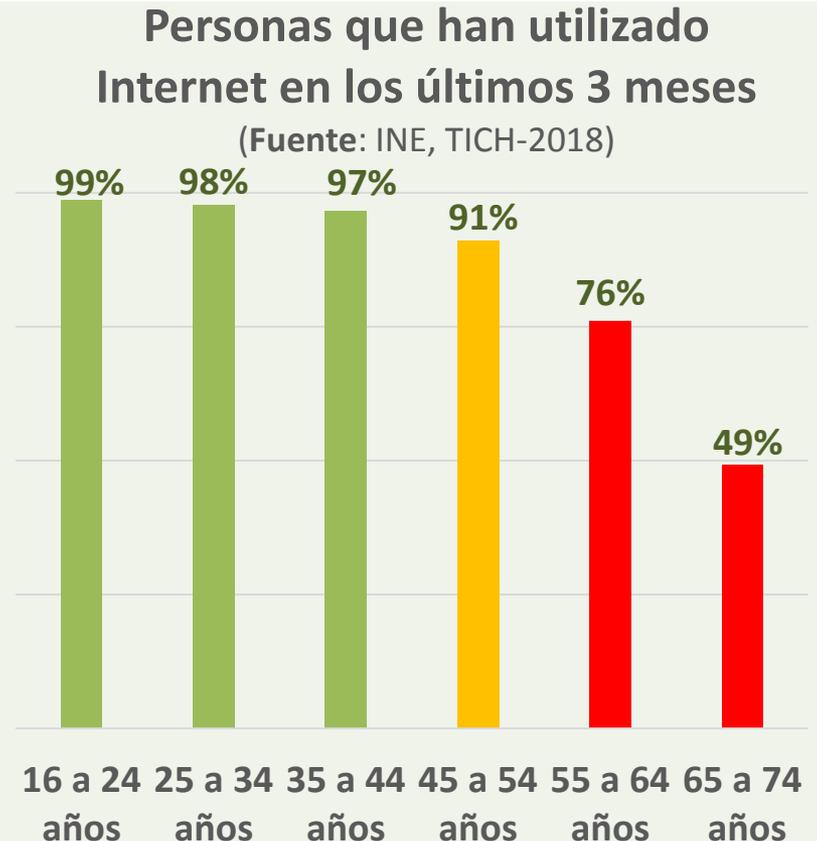
- **1973 Dympanel** implanta el 1er. panel de consumidores en España
- **1986 Telepanel** de la Universidad de Amsterdam
- **2001 INE** Realización de encuestas a través de internet: Censo -> Sistema ARCE -> Sistema IRIA -> Encuesta ECV (Modo mixto con panel **rotante** o pseudopanel)
- **2003 netquest.es** 1er. panel no probabilístico en España
- **2007 lissdata.nl** 1er. panel probabilístico LISS panel (Longitudinal Internet Studies for the Social sciences)
- **2007 Nielsen Online** - Nielsen/NetRatings medición de audiencias online
- **2009 Norma ISO 26362** específica de paneles
- **2010 comScore** referente de audiencias online en España
- **2013 Q Panel** Asociación de Investigación de Medios de Comunicación (AIMC)
- **2013 E-mop** Panel probabilístico de captación telefónica IMOP Insights
- **2014 PACIS** Panel Ciudadano para la Investigación Social en Andalucía (IESA-CSIC)
- **2015** Panel de Hogares CNMC probabilística, modo mixto: postal 75% Web 25%

## TIPOS DE PANELES ONLINE (i)

### No probabilísticos (paneles de acceso)

Los miembros del panel se reclutan a través de Internet, existiendo dos formas:

- Sin ningún tipo de restricción:**  
Formulario de auto-registro.  
p.e. [sondea.com](http://sondea.com)
- Invitación previa** o captación activa. p.e. [netquest.com](http://netquest.com)



## TIPOS DE PANELES ONLINE (ii)

### Probabilísticos

Los miembros del panel se reclutan a través de modos de administración que permiten seguir procedimientos de muestreo de tipo probabilístico, en general la captación es por encuesta postal, telefónica o presencial.

#### Internacional

- UK - Innovation Panel
- NL - LISS
- FR - ELIPSS
- DE - German Internet Panel
- US - American Trends Panel

#### España

- PACIS del IESA-CSIC
- E-Mop de IMOP Insights

## CALIDAD (i)

**La calidad de una encuesta a partir de un panel online se basará en dos aspectos:**

- Los errores cometidos del modo Web: **muestreo, cobertura, no respuesta y medida**
- Tipo de panel **probabilístico (↑calidad) o no (↓calidad)**

Para tratar de mejorar la calidad haciendo que la muestra seleccionada se parezca a la población en determinados parámetros, sobre todo en paneles no probabilísticos, son controlados en el momento de selección de la muestra (**estratificación**) o, a posteriori, aplicando técnicas de ponderación de datos (**post-estratificación**)

## CALIDAD (ii)

Si el investigador/a no va a crear un panel, sino que va usar el de una empresa tiene como elementos de control de calidad:

- Que la empresa tenga implementada la normas UNE ISO-20252, ICC/ESOMAR de investigación mediante encuestas.
- Que la empresa tenga implementada la norma ISO 26362 de paneles online.
- guía ESOMAR28 en la que se plantean veintiocho preguntas y respuestas para ayudar al investigador que se plantea adquirir muestras de este tipo de muestra.
- Que la empresa tenga implementada la norma UNE 139803 y WCAG 1.0 sobre accesibilidad y usabilidad Web.

## Reclutamiento de hogares para panel probabilístico

1. Método de captación
2. Marcos para la captación
3. Proceso de captación
4. Tamaño muestral
5. Diseño muestral
6. Selección de secciones
7. Selección de hogares
8. El catastro como marco muestral
9. Cobertura y fiabilidad
10. Indicadores de calidad

# Método de captación

- Captación telefónica

Ventajas: coste y tiempo

Inconvenientes:

- Marco (especialmente en niveles inferiores al estatal).
- Menor aceptación (desconfianza, dificultad para explicarlo)
- No hay posibilidad de tener aceptación firmada, sí grabación

- Captación presencial

Ventajas: Mayor posibilidad de aceptación

Inconvenientes:

- Coste y tiempo
- Dificultades con el marco

Se pueden aprovechar las encuestas puntuales que se realizan para hacer la captación (IMOP)

# Marco para la captación telefónica

- Listín telefónico

Baja tasa de cobertura (menor del 50%)

- RDD de teléfonos fijos

Faltan hogares que no tienen fijos

Baja tasa de cobertura en determinadas edades

- RDD de teléfonos móviles

Muy recomendable en paneles nacionales

En el marco no se distinguen provincias, muy caro para ámbitos más pequeños

# Marco para la captación

- Padrón municipal de habitantes

Se puede seleccionar a las personas concretas, incluso por algunas variables. INE, IECA, ESS

Solo para estadística oficial

- Selección de viviendas por rutas aleatorias

Habitualmente no se utilizan revisitas.

- No hay una absoluta seguridad de la aleatoriedad de la elección
- Es muy complicado el controlar cuales son las viviendas que deben ser elegidas, qué intentos ha realizado el encuestador, controles presenciales...
- No puede haber fase previa de aviso

# Marco para la captación

- Alternativa: Catastro

Ventajas:

- Listado exhaustivo de viviendas
- Disponibles la dirección y localización
- Accesible de forma libre y gratuita
- Se podría hacer envío postal como aviso, pero...

Desventajas

- Sin datos de las personas residentes, hay viviendas vacías o segundas residencias
- Las direcciones en muchos casos no se corresponden con las direcciones postales

Parece que el IECA tiene avanzada una base de datos de viviendas en Andalucía, pero aún no está disponible

# Proceso de captación

- 1<sup>a</sup> Fase: Envío postal de tríptico informativo del panel y carta de invitación para participar en el mismo. En dicha carta se ofertarán distintas vías para darse de alta en el mismo, acceso online y llamada a línea 900.
  
- 2<sup>a</sup> Fase: Recuerdo postal a las direcciones válidas que no hayan contactado .  
En las cartas devueltas, búsqueda de direcciones alternativas y nuevo envío
  
- 3<sup>a</sup> Fase: Captación personalizada. Encuestadores se encargarán de contactar con los hogares incluidos en la muestra que no hayan contactado.  
Los encuestadores llevan las direcciones de cada vivienda seleccionada en la muestra, y un plano con su localización.

# Proceso de captación

## Formulario WEB para el acceso on-line

### PACIS Panel Ciudadano para la Investigación Social en Andalucía



Área de hogar panelista ([Cerrar sesión](#))

Ficha 0 % rellena

Atención al panelista - 900 80 92 32 | 957 76 05 35 | vía [email/Web](#)

**Muchas gracias por acceder** a la ficha de inscripción en el Panel Ciudadano para la Investigación Social en Andalucía. En él pueden participar todas las personas que convivan en el hogar y que tengan 16 años o más. **Su opinión es muy importante y permitirá una mejor comprensión de nuestra sociedad andaluza.**

A continuación le haremos algunas preguntas que nos permitirán clasificar al hogar y sus miembros y contactarles para que respondan al cuestionario de los estudios que se pongan en marcha en el futuro. Le recordamos que la confidencialidad de sus respuestas está protegida por la ley 12/89 y la ley orgánica 15/99 como exponemos en el apartado de [privacidad](#).

¿Podría indicar si disponen en su vivienda de los siguientes equipamientos?

	Uno	Más de uno	No hay
Teléfono fijo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teléfono móvil	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ordenador (PC, Portátil, Tablet)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conexión Internet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Pulse aquí para continuar](#)



© 2014 | IESA-CSIC - C/ Campo Santo de los Mártires, 7 - 14004 Córdoba (España)

# Tamaño muestral

**Tamaño muestral final:** Debemos decidir cual queremos que sea el tamaño de nuestro panel, teniendo en cuenta las encuestas que vamos a hacer, la tasa de respuesta en cada una de las olas, y la previsión de pérdida de individuos en el panel

Según esa cifra, debemos calcular cuantas direcciones seleccionar, haciendo alguna estimación de la tasa de captación

# Tamaño muestral

## Ejemplo PACIS:

Nivel de error de las encuestas del PACIS:

Error típico muestral máximo objetivo:  $\pm 3\%$  (95% de confianza)

- Según m.a.s.: 1.056 entrevistas
- Tasa de respuesta prevista en cada oleada: 50%

Teniendo en cuenta ambos factores, estimamos la necesidad de tener al menos 2.400 hogares en el panel.

Suponemos:

- Un 20% de las direcciones van a corresponder a viviendas vacías o inexistentes
- Tasa de colaboración: Entre el 40% y 50% de los hogares
- De estos supuestos llegamos a la conclusión de que necesitamos seleccionar a 10.000 viviendas como muestra inicial.

En todo muestreo en hogares es inviable el muestreo aleatorio simple

- Muestreo bietápico
- Fase 1. Selección de sección censal (y núcleo)
  - Muestreo estratificado por provincias y otras variables (nivel de estudios, comportamiento electoral, etc...)
- Fase 2. Selección de viviendas en la sección
  - Cálculo del número de viviendas previstas y seleccionadas por sección. **En el PACIS**, seleccionamos 400 secciones
  - En la vivienda se intenta captar a todos los componentes de la vivienda

# Selección de secciones

Se debe conseguir una selección de hogares proporcional al conjunto de hogares del universo. Primero debemos seleccionar las secciones y después los hogares en ellas.

Selección de las secciones con probabilidad proporcional:

- Proporcional al nº de hogares: inviable (censo de 2011 insuficiente)
- Proporcional al nº de individuos en Padrón
  
- Selección de núcleos en las secciones en donde hay más de uno. Se desechan los que no cumplan algunas condiciones mínimas. Se realiza selección aleatoria de los restantes

# Selección de hogares

Tener en cuenta:

- nº de elementos en la muestra final mínima para que haya representatividad por sección.
- Dificultad particular de cada sección

## En el PACIS

- El objetivo es llegar a 6 elementos en la muestra final por sección (mínimo 4).
- Hay secciones donde va a ser más difícil obtenerlo, por:
  - Estructura urbanística
  - Direcciones mal recogidas
  - Elevado número de viviendas secundarias o vacías

Partiendo de 19 como nº mínimo, aumentamos el nº de cartas según esos parámetros.

# El Catastro como marco muestral

Accesible en <https://www.sedecatastro.gob.es> pero descarga por municipio  
Se puede pedir la base de datos entera.

Dos conjuntos: alfanumérico (fichero .cat) y cartográfico (fichero .shp)

Ficheros separados para urbana y rústica

Fichero .cat: Fichero importable para Excel con toda la información de varias tablas.

De todas ellas, utilizaremos solo una: la tabla 15: registro de inmueble

# El Catastro como marco muestral

## Asignación de sección, núcleo y código postal

No es viable utilizar el callejero del INE ni el CDAU por no concordar los nombres de las direcciones en muchas viviendas, ni presentar la dirección catastral

Asignación por ubicación espacial

Secciones: seccionado INE del Censo de 2011 (descargable en [www.ine.es](http://www.ine.es))

Núcleo de población: Capa su-03 del DERA-100, y asignación manual a los núcleos del INE

Códigos postales (para envío de cartas): Cartociudad, revisado con la web de correos manualmente

# Cobertura y fiabilidad del catastro

**Nº viviendas en catastro:** 4.263.721

**Nº viviendas INE 2.001:** 4.350.095 -> cobertura superior al 95%

## Duplicidad

Viviendas que son más de un inmueble según catastro, o inmuebles con más de una vivienda. No creemos que sea numéricamente muy importante.

**Sobrecobertura:** No altera la fiabilidad estadística, pero es una pérdida de tiempo y dinero

- Viviendas no construidas, o no entregadas.
- Uso de vivienda irreal: Por ejemplo, uso de viviendas como oficinas.

**Falta de cobertura:** Elementos que se quedan fuera del marco.

Secciones con viviendas irregulares y cuevas. Más personas empadronadas que viviendas (en las secciones con viviendas irregulares también falla el censo) Localmente importante, pero no demasiado en el global

## Problemas de identificación:

Direcciones postales. Los nombres de catastro están mal o anticuados en un gran número de casos

Recurrimos a CDAU, listines telefónicos y callejeros del INE, y sobre todo al mapa de localización



# Indicadores de caliadd

Cálculo de indicadores de calidad:

- Tasa de contacto: N° de viviendas alcanzadas respecto al total de viviendas que se incluyen en la muestra inicial.
- Tasa de cooperación: N° de hogares que aceptan incorporarse al panel con respecto a los hogares contactados.

Tasa final de respuesta (RR3): proporción de hogares captados dividido por el número de hogares contactados, más los hogares de elegibilidad desconocida corregidos estos últimos por un factor de elegibilidad. En nuestro caso el factor los calculamos de la propia muestra, asumiendo que el % de elegibles entre los de elegibilidad desconocida será el mismo que el de los de elegibilidad conocida.

## Resultados PACIS

Tasa de contacto del 64,3% (se esperaba 80%). Debido a los problemas con las direcciones, y a muchas viviendas vacías (también segundas residencias)

Tasa de cooperación: 32,3% (se esperaba 40-50%)

Consecuencia: No se estaba cumpliendo el mínimo de panelistas por sección. Se tuvo que optar por una solución a mitad de campo: aumentar los contactos, intentando contactar con las viviendas contiguas de las que estaban en la muestra.

Como resultado, de 10000 viviendas en la muestra inicial pasamos a 17.517

# Indicadores de calidad

## Resultados finales (hogares):

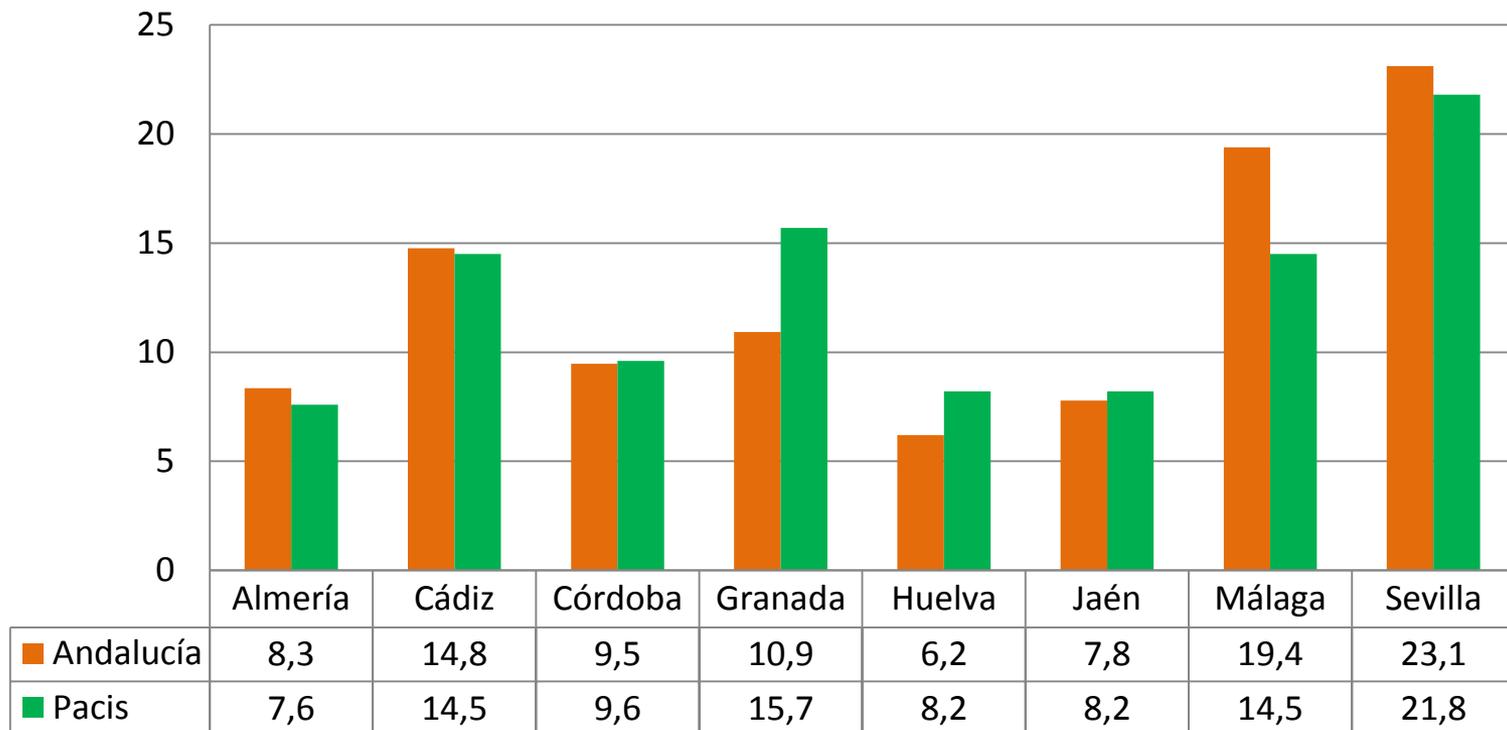
<b>Total de unidades para las que se ha intentado el contacto</b>	<b>17.517</b>
<b>Realizadas</b>	<b>2.803</b>
En la fase de envío de cartas (por web o línea 900)	186
En la fase presencial	2.617
<b>Elegible, sin realizarse</b>	<b>5.886</b>
Rechazo	5.474
Se contacta pero se acuerda nueva visita	378
Problemas de lenguaje	34
<b>Elegibilidad desconocida</b>	<b>6.123</b>
No hay nadie	5.690
No se puede acceder a la vivienda	308
No se localiza la vivienda	125
<b>No elegible</b>	<b>2.345</b>
No hay vivienda o está inhabitable	866
Se confirma que no vive nadie	1.479

**Tasa final de respuesta (RR3): 20,7%**

# Evaluación de la representatividad del panel

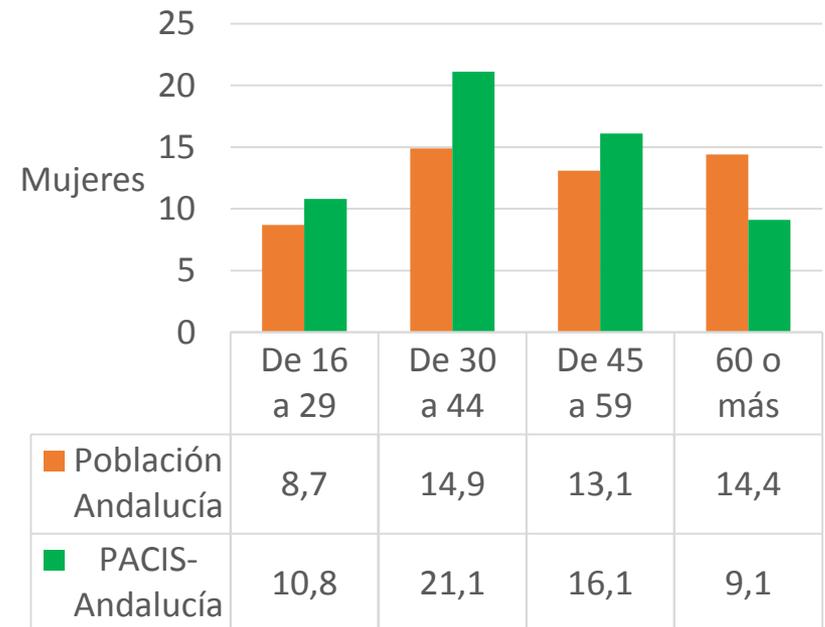
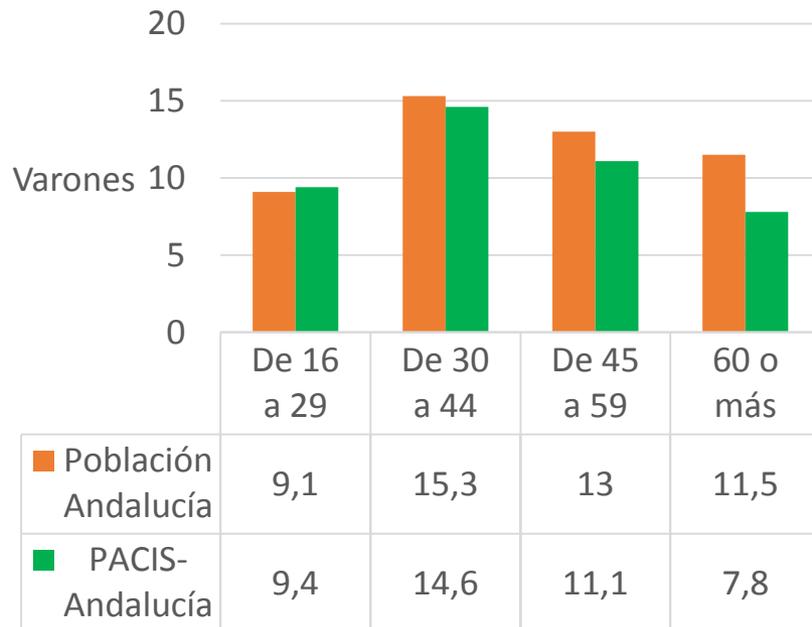
# Análisis de representatividad

- Distribución del panel comparando con datos disponibles de la población objetivo. Padrón municipal
- Provincias. **En el PACIS**



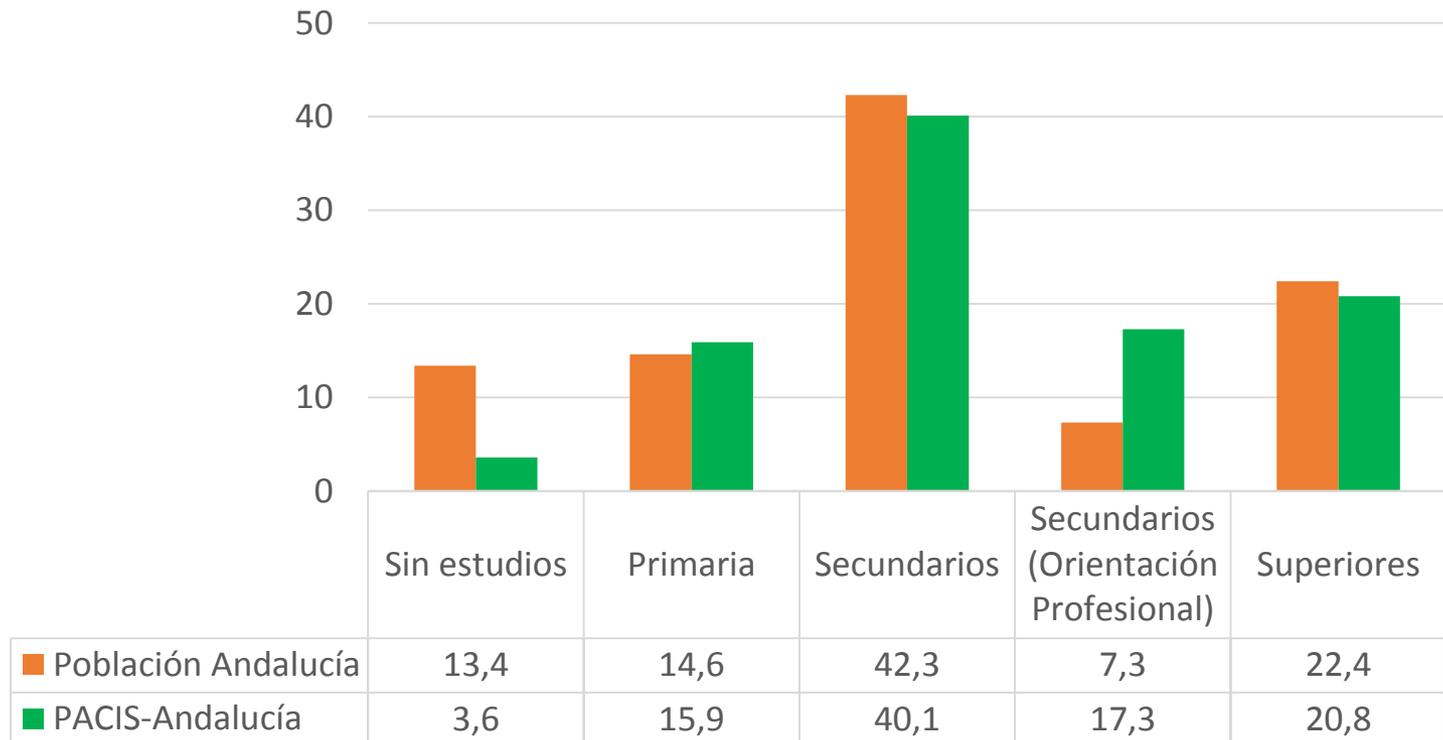
# Análisis de representatividad

- Sexo y edad. **En el PACIS**



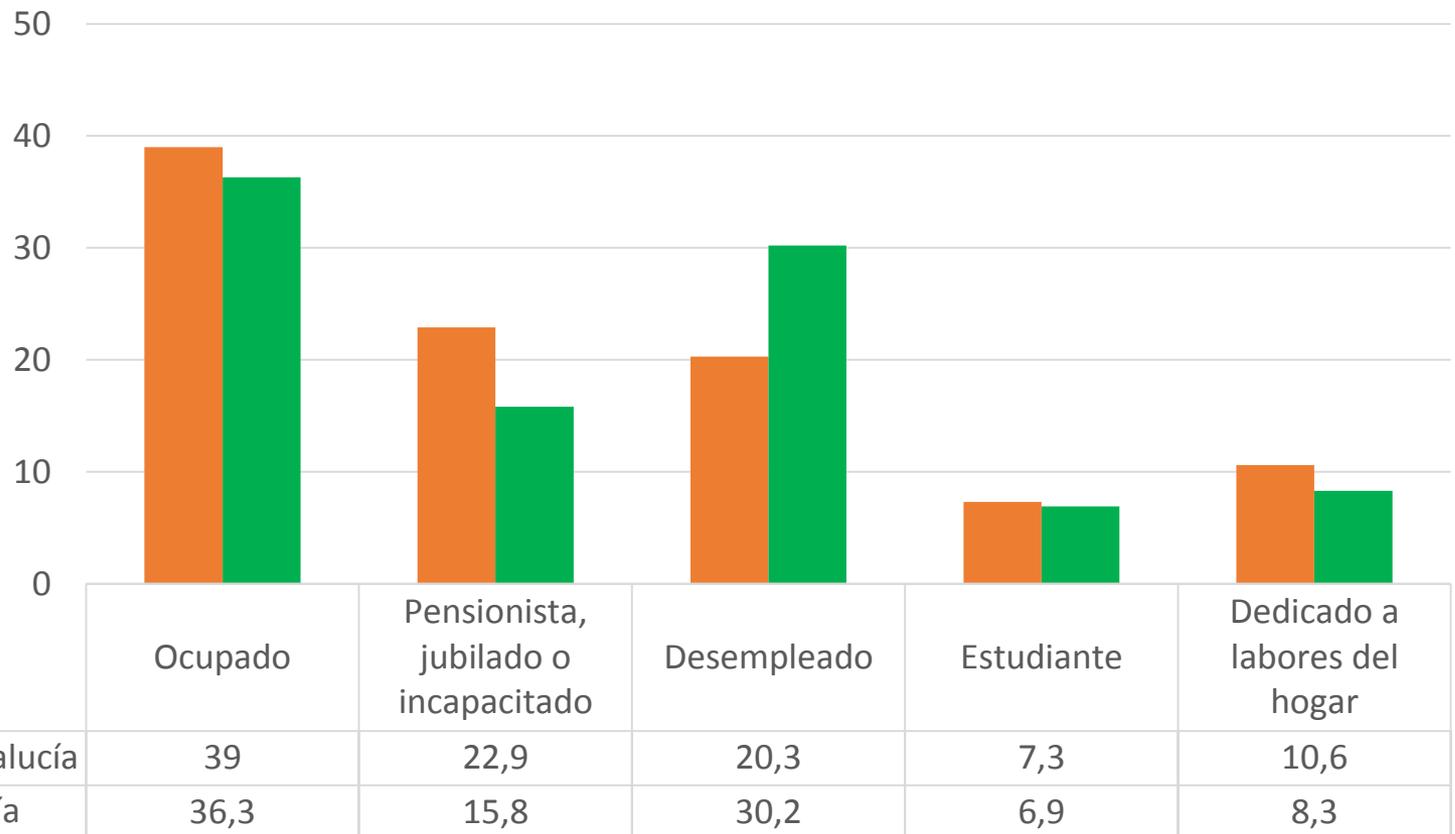
# Análisis de representatividad

- Nivel de estudios. **En el PACIS**



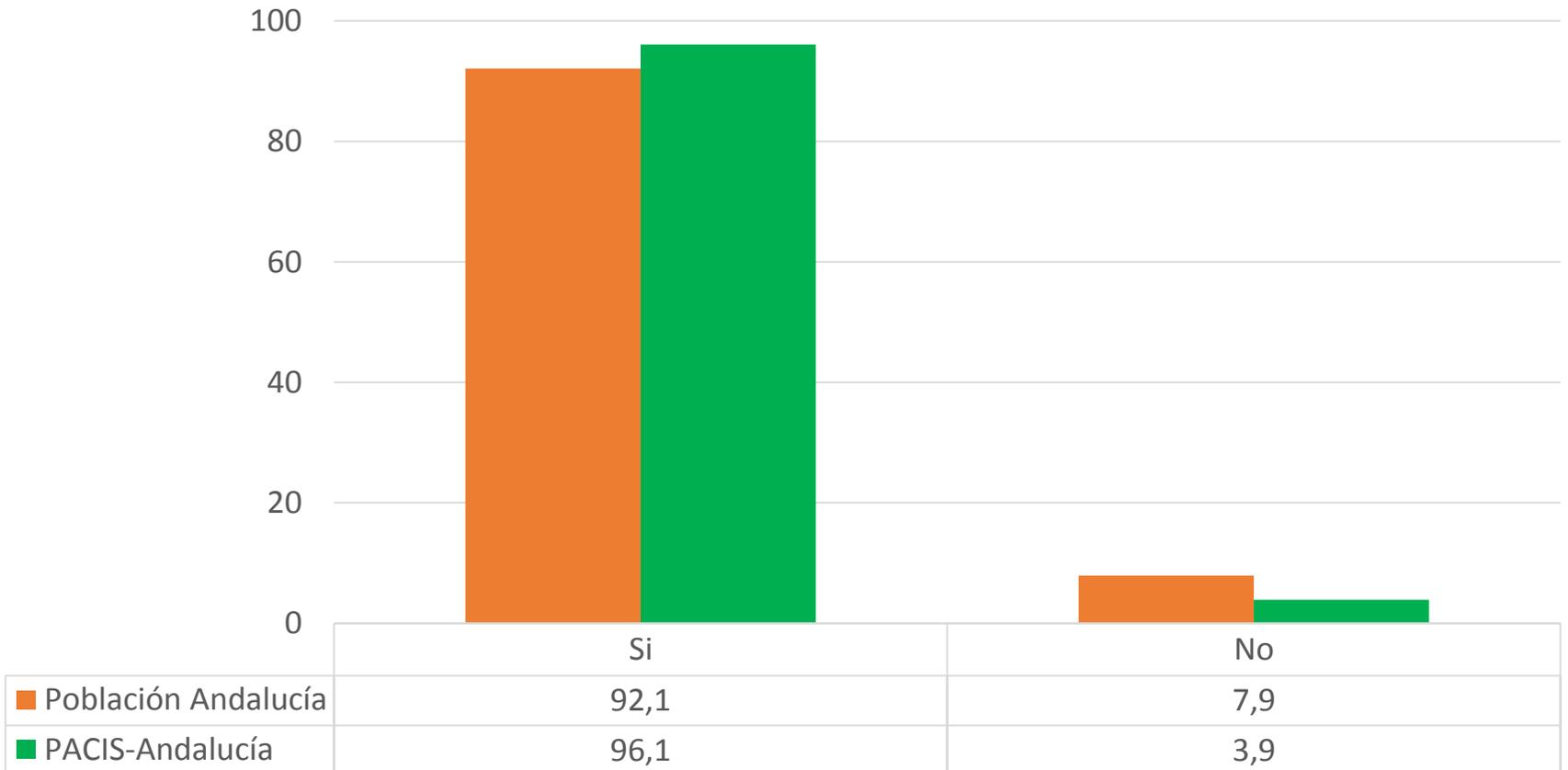
# Análisis de representatividad

- Relación con la actividad. **En el PACIS**



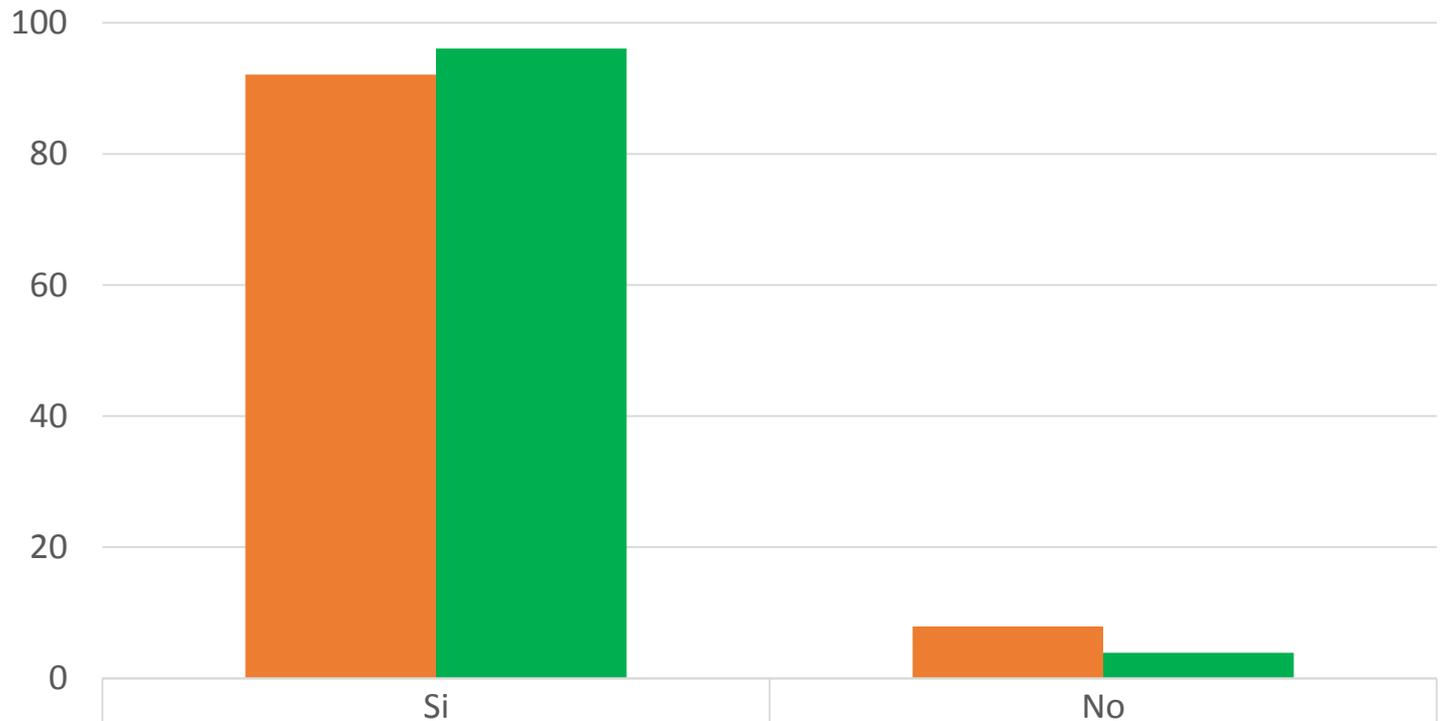
# Análisis de representatividad

## ■ Nacionalidad. **En el PACIS**



# Análisis de representatividad

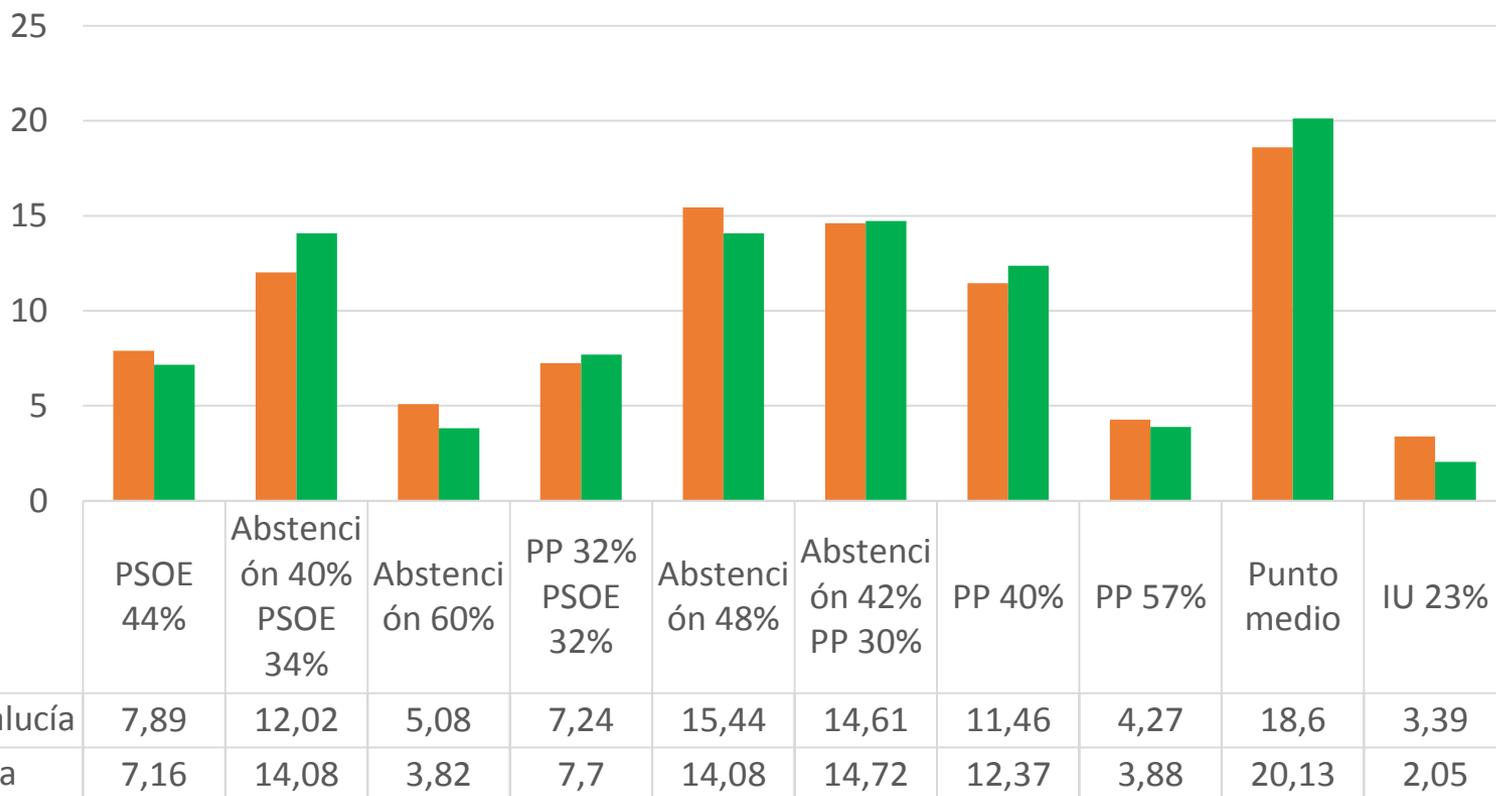
- Nacionalidad. **En el PACIS**



	Si	No
■ Población Andalucía	92,1	7,9
■ PACIS-Andalucía	96,1	3,9

# Análisis de representatividad

- Distribución por comportamiento político de la sección (en la fase de captación no se preguntó el voto). **En el PACIS**



# Realización de las encuestas

1. Diseño muestral
2. Ejecución del trabajo de campo
3. Análisis de la no respuesta

# Diseño muestral

- Muestreo estratificado por sexo y grupo de edad. Se podrían incluir otras variables que estén en la ficha de hogar.
- Afijación proporcional a la tasa de respuesta, para conseguir muestra aproximadamente equilibrada
- Inclusión solo de una persona por hogar
- Calibrado de los datos obtenidos

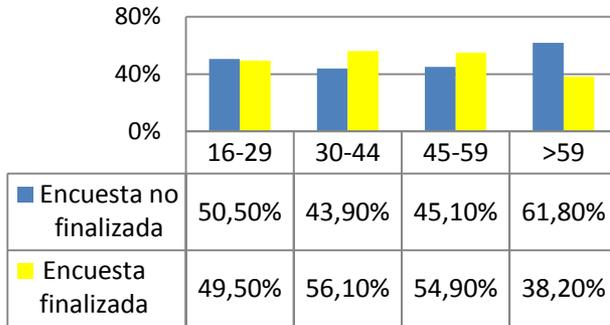
# Ejecución del trabajo de campo

1. Envío de carta postal con clave que permite acceder a la encuesta
2. Envío de correo electrónico con el enlace
3. Envío de sms facilitando la clave para hacer la encuesta
4. Entre 3 y 4 recuerdos por correos electrónico y 1 por sms. Deben ser estudiados. Entre 2 y 3 semanas
5. Las personas que no han rellenado la encuesta por CAWI más las que no tienen correo electrónico pasan al CATI.

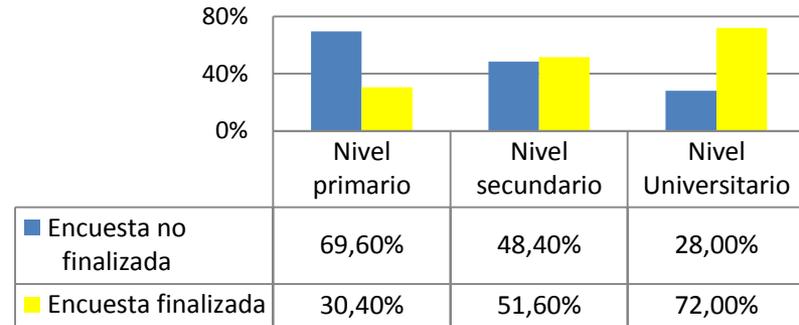
Es una gran ventaja si las interfaces de CAWI y CATI son las mismas, para permitir retomar encuestas sin finalizar en CAWI

# Análisis de la no respuesta

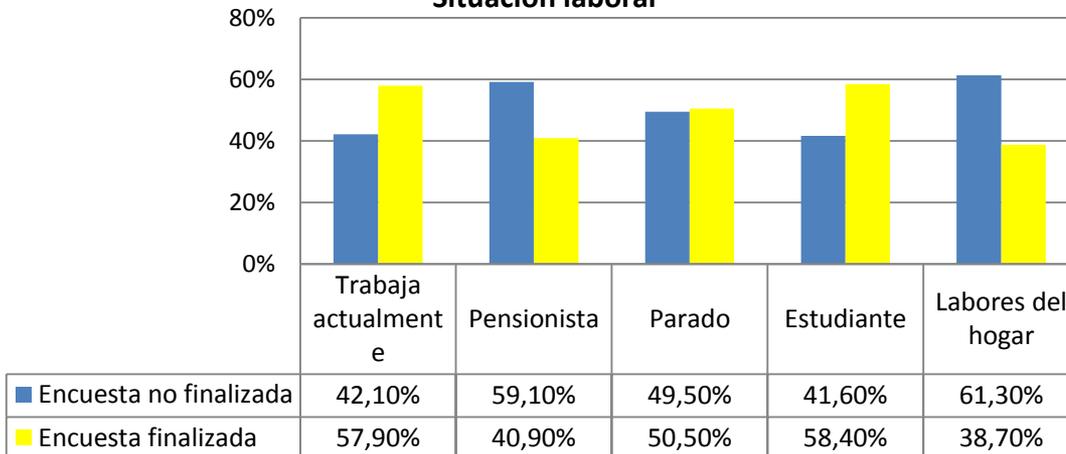
### Grupos de edad



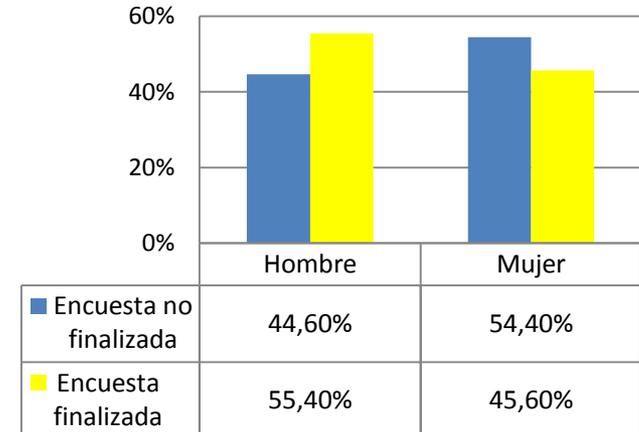
### Nivel de estudios



### Situación laboral

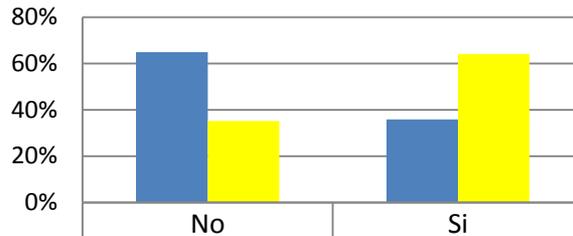


### Sexo



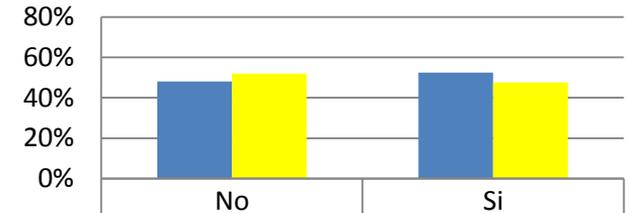
# Análisis de la no respuesta

**Posesión de email**



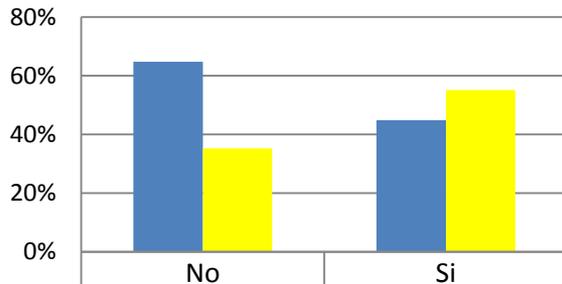
	No	Si
Encuesta no finalizada	64,90%	35,90%
Encuesta finalizada	35,10%	64,10%

**Posesión de teléfono fijo**



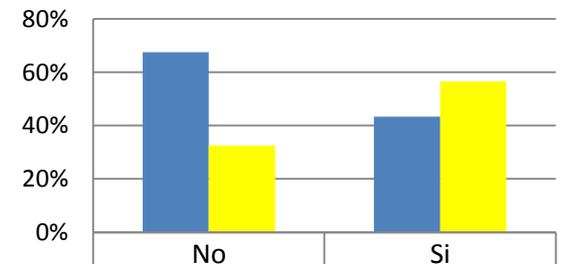
	No	Si
Encuesta no finalizada	48,10%	52,50%
Encuesta finalizada	51,90%	47,50%

**Posesión de teléfono móvil**



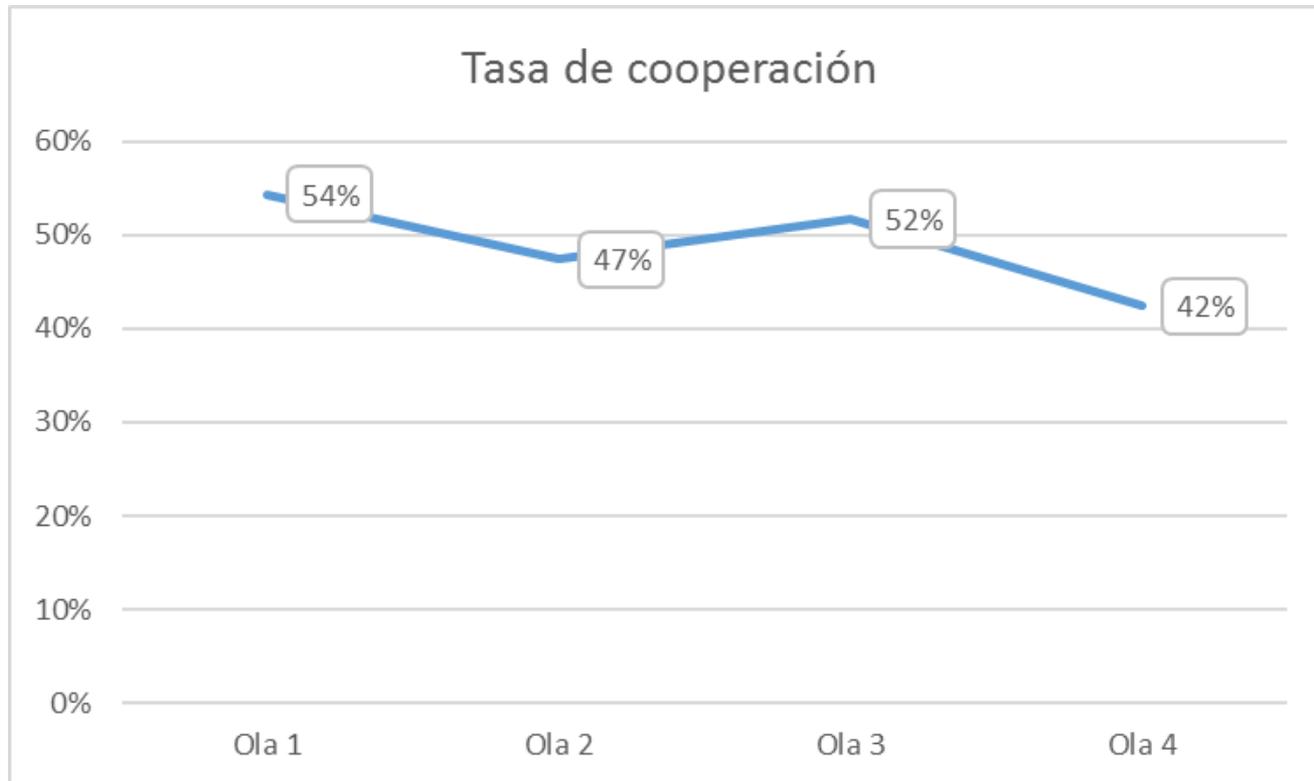
	No	Si
Encuesta no finalizada	64,80%	44,90%
Encuesta finalizada	35,20%	55,10%

**Posesión de Internet**



	No	Si
Encuesta no finalizada	67,50%	43,40%
Encuesta finalizada	32,50%	56,60%

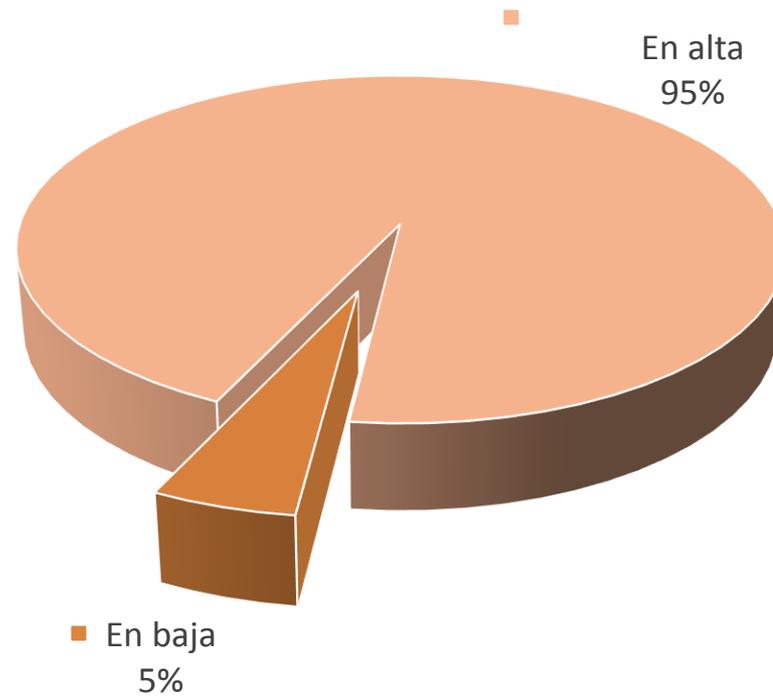
# Tasa de cooperación



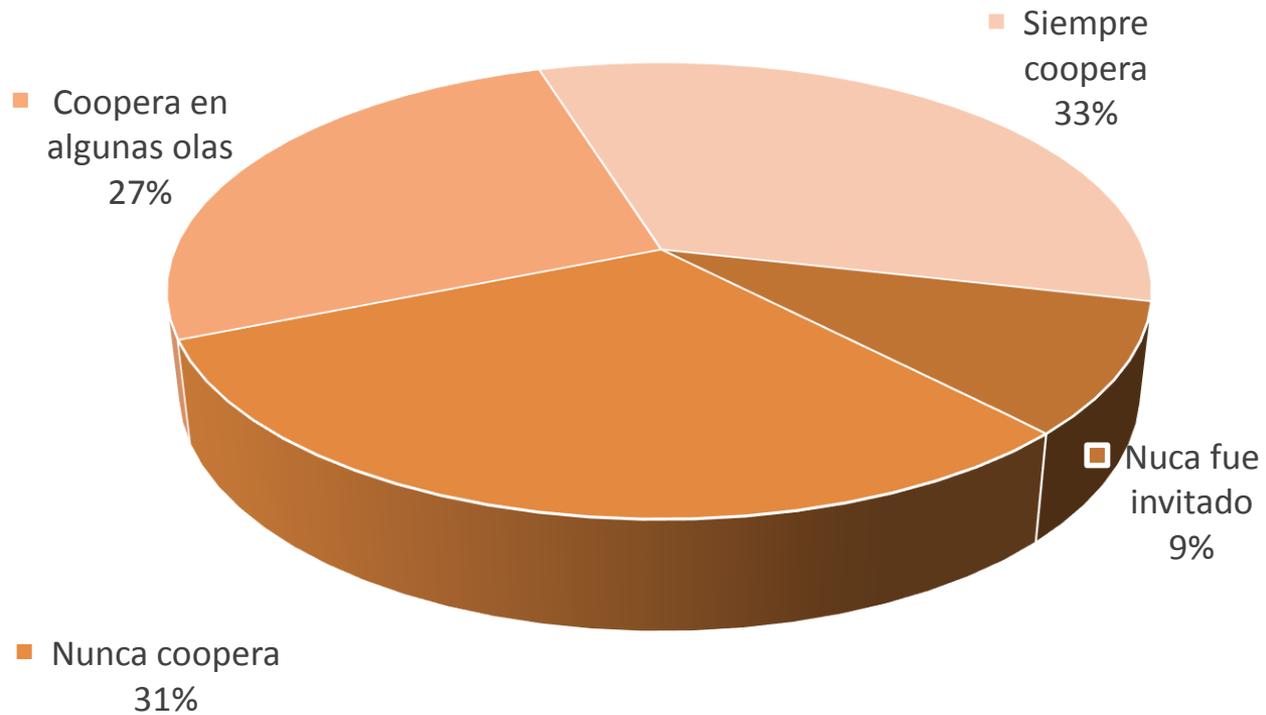
## Mantenimiento del panel

1. Bajas del panel
2. Patrones de cooperación
3. Refresco del panel

# Bajas del panel



# Patrones de cooperación



# Refresco del panel

Necesidad de refresco del panel:

- Bajas
- Tasas de cooperación se reducen

## Análisis de viabilidad

	Grupo de edad	Nº encuestas para N=1500	Marco 2017	Tasas de respuesta	Encuestas válidas estimadas	déficit	Muestra necesaria
Hombres	De 18 a 29	137	261	45,7%	119	18	39
Hombres	De 30 a 44	230	483	55,2%	267	-37	
Hombres	De 45 a 59	194	408	57,8%	236	-42	
Hombres	60 o más	172	306	47,1%	144	28	59
Mujeres	De 18 a 29	131	262	48,1%	126	5	10
Mujeres	De 30 a 44	223	697	51,4%	358	-135	
Mujeres	De 45 a 59	197	616	46,2%	285	-88	
Mujeres	60 o más	216	363	32,3%	117	99	306
		1500	3396				

Decisión: refresco del panel centrado en determinados grupos de edad

IX Congreso Andaluz de Sociología  
Huelva 22 de Noviembre

**¡¡¡Gracias!!!**



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE CIENCIA, INNOVACIÓN  
Y UNIVERSIDADES



**CSIC**

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

**IESA** Instituto de Estudios  
Sociales Avanzados

