



# UNIVERSIDAD DE GRANADA

**TRABAJO FINAL DE INVESTIGACIÓN  
MÁSTER EN ESTADÍSTICA APLICADA**

*Técnicas de estimación indirecta: aplicación a una encuesta sobre  
el cumplimiento de las reglas en la pandemia.*

**Autor: Jaime Casado Chica.**

**Tutoras: María del Mar Rueda y Beatriz Cobo.**

**Curso: 2020/2021**

## **AGRADECIMIENTOS**

No podría terminar este trabajo sin agradecer profundamente a todas esas personas que me han ayudado de diferentes formas y que me han dado todo lo necesario para poder llevar a cabo esta oportunidad.

Primero a mis tutoras Beatriz y María del Mar. Porque aún llegando con el tiempo justo, aceptaron que hiciera este trabajo y me guiaron en todo momento, dándome conocimientos, consejos y todas las respuestas que necesitaba.

A Shaila que me ha aguantado y animado durante todo este tiempo.

A mis padres que siempre hacen posible que pueda abrir puertas, y no tener miedo de cerrar otras. El resto de la familia también.

A mis amigas y amigos por esos momentos de desahogo y recargar baterías.

A toda esa gente que se dedica a la gran labor de crear, mantener y contribuir en el software libre, como R.

A la ciudad de Barcelona, y a todos los que tengo aquí, por ayudarme a crecer.

## RESUMEN

Las tradicionales encuestas tienen sus limitaciones cuando se pregunta de forma directa por algún tema cargado de sensibilidad. Esto provoca desconfianza por parte de los encuestados, lo que hace que al final la información sea errónea o incompleta.

Cuando entra en el juego el factor de la “deseabilidad social” (el comportamiento que se espera de nosotros como individuos sociales, un comportamiento calificado como correcto) se provoca un sesgo en las respuestas dadas y por tanto, unas estimaciones en las prevalencias de las conductas que estemos midiendo incorrectas.

Aparte de la respuesta sesgada, un problema grave es la participación. Cuando los encuestados perciben el cuestionario como herramienta que bien puede servir para ser juzgados por conductas comprometidas, pueden eludir, no sólo responder preguntas, sino la total participación en el estudio.

Las técnicas de estimación indirecta tratan de atajar este tipo de problemas mediante diferentes métodos y técnicas. Abarcan desde las técnicas respuestas aleatorizadas, hasta las de respuestas transversales (o cruzadas), pasando por técnicas de respuesta forzada.

Las técnicas de respuesta aleatorizada utilizan objetos externos como cartas o dados para hacer que el encuestado responda dependiendo de los resultados obtenidos en alguno de estos juegos. En las de respuesta forzada el encuestado responde en base a una respuesta dada por una instrucción que, de alguna manera, elige. Las técnicas transversales (traducción de Crosswise, el método que destacamos y usamos en este estudio) se valen de un procedimiento en el que hacen dos preguntas a la vez, obteniendo una respuesta que contemple lo que se contesta en esas dos preguntas. Es decir, una opción sería afirmar las dos preguntas o negarlas, y otra, negar una de ellas y afirmar otra.

Diferentes técnicas más destacables son explicadas en este trabajo y se profundiza en el Método Crosswise, con el que también se lleva a cabo el diseño análisis de una encuesta real aplicada, relacionada con las normas de prevención de la pandemia de Coronavirus.

En la parte práctica se han aplicado dos encuestas, una directa y otra indirecta diseñada en base al Método Crosswise. Posteriormente se hace análisis y contraste de las estimaciones de prevalencia para discutir la efectividad de las técnicas indirectas en la reducción del sesgo de deseabilidad social.

Por último, se discute el cumplimiento de la población sobre las medidas para frenar la Covid-19, fenómeno de gran importancia y preocupación en nuestro presente.

## **ABSTRACT**

The traditional surveys have their limitations when asking directly about a sensitive topic. This causes distrust on the part of the respondents, which, in the end, makes the information erroneous or incomplete.

When the factor of "social desirability" comes into play (the behavior that is expected of us as social individuals, a behavior qualified as correct), a bias is caused in the answers given and therefore, estimates in the prevalence of the behaviors that we are measuring are incorrect.

Apart from the biased answer, a serious problem is participation. When respondents perceive the questionnaire as a tool that can well serve to be judged for committed behaviors, they can avoid not only answering questions, but also total participation in the study.

Indirect estimation techniques try to tackle this type of problem using different methods and techniques. They range from randomized response techniques, to cross-sectional (or crossed) responses, through forced response techniques.

Randomized response techniques use external objects such as cards or dice to make the respondent respond depending on the results obtained in one of these games. In forced response, the respondent responds based on a response given by an instruction that, in some way, he chooses. The transversal techniques (Crosswise translation, the method that we highlight and use in this study) use a procedure in which they ask two questions at the same time, obtaining an answer that contemplates what is answered in those two questions. That is, one option would be to affirm the two questions or deny them, and another, to deny one of them and affirm another.

Different most notable techniques are explained in this work and the Crosswise Method is explored, with which the analysis design of a real applied survey is also carried out, related to the norms for the prevention of the Coronavirus pandemic.

In the practical part, two surveys have been applied, one direct and the other indirect, designed based on the Crosswise Method. Subsequently, the prevalence estimates are analyzed and contrasted to discuss the effectiveness of indirect techniques in reducing the social desirability bias.

Finally, the population's compliance with the measures to stop Covid-19 is discussed. A phenomenon of great importance and concern in our present.

# Índice

<b>1. Introducción del problema.</b>	<b>1</b>
1.1. COVID-19. El surgimiento de una pandemia.	1
1.2. Las medidas contra la Covid-19.	4
1.2.1. Confinamiento Domiciliario.	5
1.3. Estudios sobre la COVID-19 con técnicas directas.	8
1.3.1. Encuestas del Centro de Investigaciones Sociológicas.	8
<b>2. Las técnicas indirectas.</b>	<b>21</b>
2.1. Modelo Warner.	22
Aplicación en un diseño muestral general.	27
2.1.2. Ampliaciones y mejoras destacables del Modelo Warner.	28
Modelo de Respuesta no relacionada.	29
Modelo C	30
Modelo de respuesta forzada.	31
Modelo de Devore	33
2.2. Modelo Crosswise.	35
2.2.1. Variación del Modelo Crosswise: el modelo triangular.	38
2.3. Regresión logística para el Modelo Crosswise.	41
2.3.1. Regresión logística en respuestas aleatorias univariantes.	41
2.3.2. Estimadores para el Modelo Crosswise.	43
2.3.3. Regresión logística para respuestas aleatorias multivariantes.	45
2.4. Ejemplo de aplicación del Método Crosswise en un estudio sobre el Coronavirus.	46
Metodología empleada.	46
Resultados.	47
<b>3. Aplicación a una encuesta real.</b>	<b>49</b>
3.1. Objetivos del estudio.	49
3.2. Universo y muestra.	50
3.3. Metodología.	51
3.4. Fases del estudio y programación.	51
3.5. Diseños de las diferentes encuestas.	53
3.5.1. Encuesta directa.	53
3.5.2. Encuesta indirecta.	56
<b>4. Análisis.</b>	<b>61</b>
4.1. Depuración de datos y pasos previos al análisis.	61
4.2. Composición de la muestra y sus características.	63
4.3. Análisis de los datos.	67
4.4. Resultados.	68
<b>5. Conclusiones.</b>	<b>73</b>
<b>6. Bibliografía.</b>	<b>75</b>
<b>ANEXOS.</b>	<b>80</b>

# 1. Introducción del problema.

## 1.1. COVID-19. El surgimiento de una pandemia.

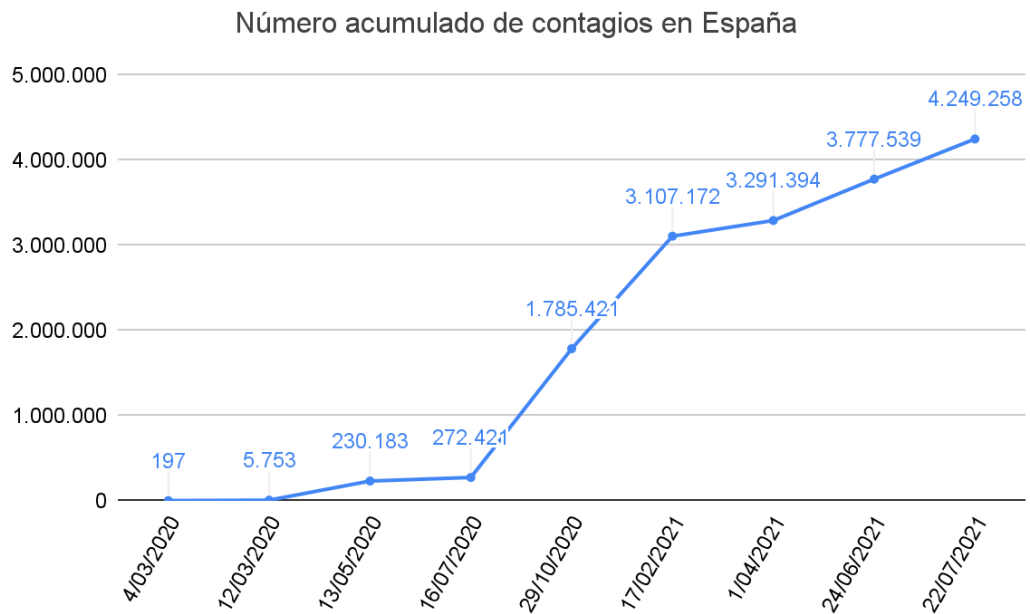
En el año 1960, en Inglaterra, ya se detectaron los primeros casos en humanos de coronavirus (Ruiz Bravo & Jimenez Valera, 2020). Una enfermedad respiratoria provocada por un virus. Pocos años más tarde, se acuñó el término “Coronavirus” para definir a este género de virus, en el que se encuentra la familia “Coronaviridae”, dentro de la cual hay dos especies concretas que afectan a los humanos. No fue hasta el año 2002 que se denominó SARS (“Severe acute respiratory syndrome”), y pasó a formar parte de un conocido grupo de enfermedades infecciosas similares a la tradicional gripe.

A finales del año 2019 en Wuhan, China, se comienzan a reportar un número elevado de casos que cumplen los criterios preestablecidos para la identificación de virus emergentes y reemergentes (Aragón Nogales, et al. 2020). Criterios basados en síntomas como fiebre, linfocitos bajos, neumonía, no mejora de los síntomas tras administración de tratamiento antimicrobiano, etc. Precedentes como las epidemias, de una magnitud menor, la “SARS-CoV” y la “MERS-CoV” definen estos criterios sintomatológicos. Esto puso sobre alerta a las autoridades sanitarias chinas, que, tras el aumento de casos, decidieron tomar medidas de control y aislamiento de la población. El 30 de enero de 2020 la Organización Mundial de la Salud declaró el virus como una emergencia a nivel mundial. El 11 de febrero de 2020 la enfermedad pasó a llamarse COVID-19, y el nombre del virus, debido a su análisis genómico, “SARS-CoV-2”. A día de hoy, se han contabilizado más de 191 millones de casos, y más de 4 millones de muertes en todo el mundo, según el seguimiento actualizado que realiza la Universidad Johns Hopkins (2021).

Contextualizados en la magnitud de esta pandemia, los datos ofrecen una evolución cronológica importante y con ciertas fechas destacables.



Se van a mostrar ahora unas líneas de tiempo con diferentes datos relevantes para analizar la importancia del problema.



Fuente: elaboración propia con datos de Statista (2021)

Se aprecia una tendencia exponencial diferenciada en lo que se denominan diferentes “olas de contagios”. La primera, abarcó el periodo de confinamiento, la segunda fue después del verano de 2020, una tercera y muy importante en las navidades 2020-2021, una cuarta, un tanto menor alrededor de marzo de 2021 y la

quinta ola está teniendo lugar desde junio de 2021.



Fuente: elaboración propia con datos de Statista (2021)

Los fallecidos suelen tener un comportamiento proporcional al número de contagiados. Se observan los mismos patrones marcados por las tendencias según las olas por las que se atraviesa.

Visualizados estos gráficos se comprende la magnitud de la pandemia, y concretamente cómo golpeó España. Por ejemplo, desde julio de 2020 a febrero de 2021 se contabilizaron casi 3 millones de contagios. Una cantidad importante para un país con una población de 47.394.223 (Instituto Nacional Estadística, 2021).

## 1.2. Las medidas contra la Covid-19.

Ya en la primera epidemia SARS-CoV en el año 2002, cuyo foco principal fue en la ciudad de Singapur, se establecen unas medidas preventivas un tanto menos restrictivas que las de la pandemia SARS-CoV-2 para evitar la transmisión entre personas. Entre estas medidas encontramos: cubrirse con pañuelos de papel mientras se estornuda o se tose; cubrirse con el codo mientras se estornuda o se tose; lavarse las manos después de haber tosido, estornudado o haber limpiado la nariz; usar jabón mientras se lavan las manos; llevar una máscara que cubra la boca; usar cubiertos de plástico para comer y desecharlos cuando se come en compañía de otras personas; evitar tocar objetos directamente con las manos que toca todo el mundo, como manivelas, botones de ascensor, etc; lavar las manos después de tocar estos objetos que toca todo el mundo (Quah & Hin-Peng, 2004).

Esta vez, aunque las medidas tienen su origen en esta primera epidemia, se establecieron unas medidas mucho más estrictas. Estas son las que fijó el Ministerio de Sanidad, Consumo y Bienestar Social (2020):

- Uso de mascarilla homologada que cubra nariz y boca fuera del domicilio. O dentro si hay personas que están contagiadas o han estado en contacto con alguna persona contagiada por Covid-19.
- Mantener una distancia de seguridad de 1,5 metros en espacios públicos (calle, parques, comercios, lugares de trabajo...)
- Lavado frecuente de manos con jabón o gel hidroalcohólico. Este último obligatorio al entrar y salir de todos los espacios cerrados públicos, como por ejemplo supermercados.
- Menos contactos. Esto supone: reuniones y espacios públicos con aforo limitado. Disminuir la frecuencia y aforo con la que se hacen encuentros con gente cercana.
- Ventilación: en lugares tanto públicos, como gimnasios, como en ámbito privado.

- Obligatoriedad de realizar cuarentena en casa, y en una habitación aislado si se convive con más personas en los casos de ser positivo o haber tenido contacto con algún positivo.

### 1.2.1. Confinamiento Domiciliario.

Ante la imparable transmisión del coronavirus a lo largo de Europa y el resto del mundo, la Organización Mundial de la Salud decidió declarar la situación como pandemia mundial (Organización Mundial de la Salud, 2020). Por este motivo, y ante la complicada situación social y económica que se preveía, el Gobierno de España declaró, el 13 de marzo de 2020, el Estado de Alarma en todo el país (Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática, 2020).

Bajo la declaración del excepcional Estado de Alarma, el Gobierno decidió también tomar como medida preventiva para evitar contagios el confinamiento domiciliario. Este consiste en la retención de toda la población en sus domicilios, pudiendo sólo salir de ellos para realizar actividades de primera necesidad, tales como; ir a comprar comida, pasear a la mascota, depositar la basura al contenedor, comprar medicamentos o ir a algún centro médico por alguna emergencia.

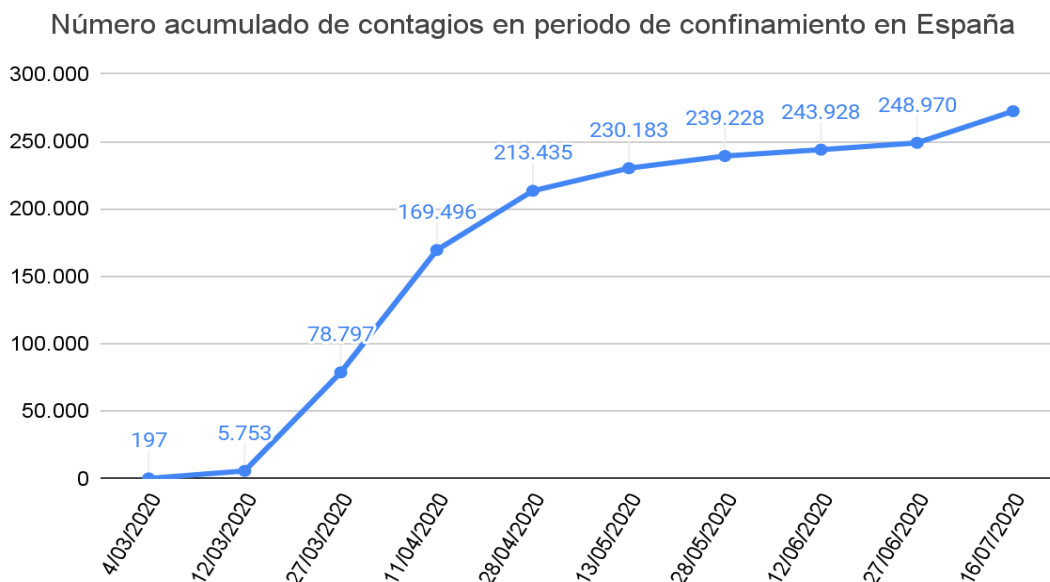
En un confinamiento domiciliario estricto no está permitido ni siquiera acudir a un trabajo que no sea denominado “esencial”. Por ejemplo, personal sanitario, cuerpos de seguridad, personal involucrado en producción y distribución de alimentación, etc.

Cabe señalar, que es una medida que genera malestar físico y psicológico, por lo que puede ser difícil para las personas cumplirla de manera estricta (Cabrera, 2020). El confinamiento es un factor clave en este estudio, pues se pretende averiguar, en qué grado la población encuestada lo llevó a cabo, y tener una prevalencia sin sesgo de deseabilidad social.

El confinamiento domiciliario estricto tuvo lugar entre las fechas 14 de marzo y 22 de junio de 2020. Aunque a partir del 2 de mayo se empezó a habilitar un plan de desescalada en el que la población podía empezar a salir de casa para actividades que no sean de primera necesidad (Canal Sur, 2021).

Se muestran ahora unas gráficas que muestran la evolución de los contagios y fallecimientos, factores sobre los que se pretende mantener el control en un periodo de confinamiento.

El primer gráfico corresponde con el periodo de confinamiento domiciliario. Una medida adoptada por el Gobierno español, bajo un Estado de Alarma para frenar la transmisión del virus (La Moncloa-Gobierno de España, 2020).



Fuente: elaboración propia con datos de Statista (2021)

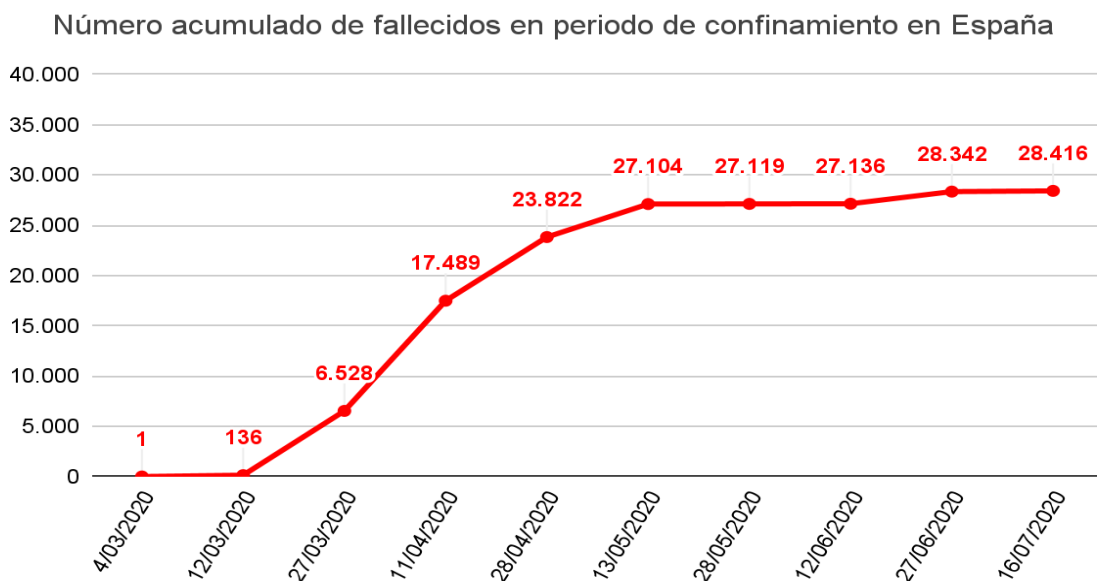
Se empiezan a contabilizar datos el 4 de marzo de 2020, tanto contagios como fallecimientos. Se selecciona esta fecha en este estudio porque es cuando se notificó el primer fallecido por Covid-19 en España (Gabinete de la Presidencia del Gobierno, 2020). Los datos concluyen un tiempo después de que se establezca “la nueva normalidad” y el fin de confinamiento domiciliario total o parcial.

Puede observarse un repunte de casos desde el inicio del confinamiento, hasta que se produce una estabilización dos meses después, el 13 de mayo, que coincide con

el fin de la curva de estabilización, es decir, cuando el aumento de casos deja de ser tan intenso y pasa a ser más distendido.

Hay que tener en cuenta que el periodo de incubación del virus, desde que un infectado lo contrae hasta que se manifiestan los síntomas, y por tanto se realiza el test, son entre 6-10 días (Rodríguez-Parrales, 2021). La notificación de resultado positivo se obtiene días más tarde, y sabremos los efectos deseados de las medidas preventivas tomadas a posteriori.

Es probable que de no haber habido un confinamiento domiciliario que limite las relaciones sociales y la movilidad, la curva antes mencionada y la estabilización no se hubieran producido.



Fuente: elaboración propia con datos de Statista (2021)

Para los fallecimientos la tendencia es similar a la de los contagios, habiendo un repunte al inicio y un cese de la intensidad de la mortalidad a fecha 13 de mayo aproximadamente.

Ante estas dos gráficas, se puede deducir que el confinamiento domiciliario fue una medida efectiva para frenar el avance del SARS-CoV-2.

### 1.3. Estudios sobre la COVID-19 con técnicas directas.

Al ser la COVID-19 una enfermedad que ha afectado al mundo entero y de forma tan intensa, la comunidad científica se ha volcado en realizar todo tipo de estudios para conocer mejor este fenómeno.

Se van a citar estudios relevantes realizados con técnicas directas de los que se pueden extraer diferentes conclusiones, y ayudarán a respaldar la hipótesis de que, el cumplimiento de las medidas impuestas por el Gobierno ha sido realmente difícil y no se ha llevado a cabo de manera estricta.

#### 1.3.1. Encuestas del Centro de Investigaciones Sociológicas.

Poco después de que empezara el Estado de Alarma por el Coronavirus, el CIS comenzó a hacer una serie de encuestas que tienen como tema central la repercusión social del fenómeno. Cómo la sociedad lo afronta en su día a día, laboralmente, doméesticamente e incluso emocionalmente.

También sirven como ejemplo de estudios con técnicas directas, aplicadas al tema que se trata en este trabajo. La metodología que se lleva a cabo en este trabajo de investigación es la técnica indirecta, en concreto el Método Crosswise, que se verá más adelante.

La encuesta nº 3302 del CIS (2020) se titula “EFECTOS Y CONSECUENCIAS DEL CORONAVIRUS (II)”. Se realizó del 23 al 26 de noviembre de 2020 y cuenta con una muestra de 2.131 entrevistas por toda España, de forma telefónica. Hay gran variedad de preguntas, pero se destacan relacionadas con este estudio y el cumplimiento de las medidas contra la Covid-19 las siguientes:

### **Pregunta 16**

*¿Cree Ud. que para que se cumplan las normas de protección contra el coronavirus, (distanciamiento social, uso de mascarillas en lugares públicos, etc.) es suficiente confiar en que las personas las cumplan voluntariamente o cree que son necesarias medidas impositivas como sanciones, multas, etc.?*

<i>Es suficiente confiar en que las personas las cumplan voluntariamente</i>	16,0
<i>Son necesarias medidas impositivas como sanciones, multas, etc.</i>	78,4
<i>N.S.</i>	4,3
<i>N.C.</i>	1,3
<b>(N)</b>	<b>(2.131)</b>

### **Pregunta 16a**

*¿Y cree que habría que endurecer más las sanciones o cree que no es necesario?*

<i>Cree que hay que endurecer más las sanciones</i>	49,4
<i>No cree que sea necesario</i>	42,2
<i>N.S./Duda</i>	7,9
<i>N.C.</i>	0,5
<b>(N)</b>	<b>(1.670)</b>



### **Pregunta 17**

¿Diría Ud. que, por lo general, se puede confiar en la mayoría de la gente, o que nunca se es lo bastante prudente en el trato con los/as demás? Por favor, sitúese en una escala de 1 a 10, en la que 1 significa 'nunca se es lo bastante prudente' y 10 significa que 'se puede confiar en la mayoría de la gente'.

1 Nunca se es lo bastante prudente	9,9
2	3,4
3	6,8
4	7,2
5	19,5
6	12,5
7	18,4
8	12,9
9	3,1
10 Se puede confiar en la mayoría de la gente	4,7
N.S.	1,1
N.C	0,5
<b>(N)</b>	<b>(2.131)</b>
Media	5,5
Desviación típica	2,4
<b>(N)</b>	<b>(2.098)</b>

Se han elegido estas preguntas cuya opinión de los encuestados respalda la complejidad de llevar a cabo las medidas de prevención contra la Covid-19.

Por un lado, en la *pregunta 16*, la mayoría de las respuestas a esa pregunta declaran no tener confianza en la voluntariedad de la gente para tomar medidas de prevención, sino que se necesita de ser un hecho coercitivo. Esto dice mucho de lo que representan las medidas para la mayoría de la población, un esfuerzo tedioso.

También en la *pregunta 16a* casi el 50% de la muestra que contestó a esta pregunta, cree que habría que endurecer las sanciones por no cumplir las medidas. Esto se traduce en que estas personas piensan que el cumplimiento no está siendo lo suficientemente estricto.

Por último, la *pregunta 17*, habla de la confianza que se puede tener hacia las otras personas. Se refiere respecto a evitar el contagio. Es decir, tener la tranquilidad de que esas personas actúan bajo las medidas de prevención de ese momento y no va a haber riesgos de contagio. Bien, en la puntuación donde 1 quiere decir que no se puede confiar y 10 que se puede confiar totalmente, la puntuación media de la pregunta, es tan solo de 5.5. La puntuación más valorada es 5. Son unos valores que no demuestran una puntuación de una confianza elevada.

De la encuesta nº 3305 del CIS (2020) que se titula “EFECTOS Y CONSECUENCIAS DEL CORONAVIRUS (III)”, y que se realizó el 11 de diciembre de 2020 y cuenta con una muestra de 2.084 entrevistas por toda España, de forma telefónica, destacan las siguientes:

### ***Pregunta 7***

*¿Usa Ud. mascarillas habitualmente como medida de protección?*

<i>Sí</i>	<i>99,2</i>
<i>No</i>	<i>0,8</i>
<i>N.C.</i>	<i>-</i>
<b><i>(N)</i></b>	<b><i>(2.084)</i></b>

### **Pregunta 7a**

*¿Me podría decir cuántas mascarillas suele Ud. usar a la semana?*

<i>Una</i>	6,4
<i>Dos</i>	9,4
<i>Tres</i>	14,3
<i>Cuatro</i>	10,8
<i>Cinco</i>	7,1
<i>Seis</i>	3,8
<i>Siete</i>	15,5
<i>Ocho</i>	2,6
<i>Nueve o más</i>	19,9
<i>Utiliza siempre la misma (no es desechable)</i>	6,3
<i>N.S., duda</i>	3,7
<i>N.C.</i>	0,2
<b>(N)</b>	<b>(2.068)</b>
<i>Media (de los que usan mascarillas desechables)</i>	5,3
<i>Desviación típica</i>	2,7
<b>(N)</b>	<b>(1.856)</b>

### **Pregunta 7b**

*Y piense ahora en un día de actividad normal. ¿Cuántas horas suele Ud. usar la mascarilla a lo largo del día, en lugares públicos (calle, comercio, centro de estudios, transporte, etc.) o de trabajo?*

<i>Menos de una hora</i>	8,4
<i>Entre 1 y 2 horas</i>	20,3
<i>Entre 2 y 3 horas</i>	13,8
<i>Entre 3 y 4 horas</i>	10,3
<i>Entre 4 y 5 horas</i>	6,8
<i>Entre 5 y 6 horas</i>	5,9
<i>Entre 6 y 7 horas</i>	3,5
<i>Entre 7 y 8 horas</i>	5,5
<i>Entre 8 y 9 horas</i>	7,4
<i>Entre 9 y 10 horas</i>	7,4
<i>Más de 10 horas</i>	8,3
<i>N.S.</i>	2,0
<i>N.C.</i>	0,4
<b>(N)</b>	<b>(2.068)</b>
<i>Media</i>	4,1
<i>Desviación típica</i>	3,3
<b>(N)</b>	<b>(2.019)</b>

### **Pregunta 8**

*¿Usa Ud. gel hidroalcohólico para desinfectarse las manos?*

<i>Sí</i>	94,2
<i>No</i>	5,8
<i>N.C.</i>	-
<b>(N)</b>	<b>(2.084)</b>

### **Pregunta 8a**

*¿Cuándo usa Ud. gel hidroalcohólico para desinfectarse las manos? (RESPUESTA MÚLTIPLE. MARCAR TODAS LAS QUE DIGA LA PERSONA ENTREVISTADA).*

<i>Siempre o casi siempre</i>	42,6
<i>Cada vez que toca algún objeto (botón, pulsador, picaporte, pasamanos...)</i>	37,6
<i>Al llegar al trabajo y durante el trabajo</i>	25,1
<i>Al llegar y al salir de casa, y en casa</i>	37,3
<i>Al entrar o salir de un comercio</i>	63,4
<i>Al entrar o salir de un bar o restaurante</i>	24,3
<i>En cualquier sitio público</i>	2,8
<i>De vez en cuando, en algunas situaciones</i>	2,7
<i>Otras respuestas</i>	1,1
<i>N.C.</i>	0,2
<b>(N)</b>	<b>(1.964)</b>

### **Pregunta 9**

*Aparte de las mascarillas y el gel hidroalcohólico, de las siguientes medidas de protección frente al virus COVID-19, ¿cuáles practica Ud. normalmente?*

	<b>Sí</b>	<b>No</b>	<b>N.C.</b>	<b>(N)</b>
<i>Guarda la distancia de seguridad entre personas recomendada por las autoridades sanitarias</i>	95,1	4,1	0,8	(2.084)
<i>Se lava las manos con frecuencia en casa u otros lugares</i>	95,8	4,1	0,1	(2.084)
<i>Desinfecta productos alimenticios antes de comerlos</i>	50,7	48,6	0,7	(2.084)

### **Pregunta 10**

*¿Utiliza Ud. alguna otra medida de protección frente a la COVID-19?*

<i>Sí</i>	24,0
<i>No</i>	75,9
<i>N.C.</i>	0,1
<b>(N)</b>	<b>(2.084)</b>

### **Pregunta 10a**

*¿Podría decirme cuál o cuáles? (RESPUESTA ESPONTÁNEA). (MÁXIMO TRES RESPUESTAS).*

<i>Desinfecta la casa, objetos, lavadora o utensilios con desinfectante, alcohol, lejía, ozono</i>	<i>44,0</i>
<i>Desinfecta y limpia zapatos o se descalza o deja fuera los zapatos</i>	<i>22,2</i>
<i>Utiliza guantes, gafas, gorros en algunos lugares</i>	<i>14,8</i>
<i>Ventila en la casa y/o en el trabajo</i>	<i>12,4</i>
<i>Mantiene la distancia, evita las aglomeraciones, vigila los aforos</i>	<i>12,2</i>
<i>Reduce el contacto social y relaciones sociales</i>	<i>6,8</i>
<i>Limita las salidas, se confina en casa</i>	<i>6,4</i>
<i>Otras respuestas</i>	<i>8,6</i>
<i>N.C.</i>	<i>0,2</i>
<b>(N)</b>	<b>(500)</b>

Las preguntas 7, 7a, 7b, 8, y 8a sobre el uso de mascarilla y gel hidroalcohólico ofrecen puntuaciones muy altas cuando son preguntadas en forma explícita, con valores alrededor del 90%. Al igual que en la pregunta, las cuestiones de la *pregunta 9*, que tienen que ver con las medidas en el momento de la interacción con otras personas (excluyendo la de lavar los alimentos) tienen también puntuaciones muy altas, y son preguntadas de forma directa. Por último lo llamativo es que, además hay bastante gente que afirma, en las *preguntas 10 y 10a* que también efectuaba labores de limpieza adicionales, sin ser explícitamente medidas obligatorias del Gobierno.

Hasta ahora, en la primera encuesta se deduce un malestar emocional y desconfianza sobre el cumplimiento de las medidas de prevención adoptadas por el gobierno. En la segunda, cuando se pregunta personalmente por si se llevan a cabo las medidas, los porcentajes afirman que se cumplen en gran medida.

Para terminar con el CIS (2021), la encuesta nº 3324 que se titula “EFECTOS Y CONSECUENCIAS DEL CORONAVIRUS (IV)”, que se realizó del 14 al 29 mayo de 2021 y cuenta con una muestra de 3.008 entrevistas por toda España, de forma telefónica, contiene las preguntas que se han destacado en la última encuesta expuesta (la 3305) más actualizadas:

### **Pregunta 7**

*¿Usa Ud. mascarillas habitualmente como medida de protección?*

Sí	99,4
No	0,6
N.C.	-
<b>(N)</b>	<b>(3.008)</b>

### **Pregunta 7a**

*¿Me podría decir cuántas mascarillas suele Ud. usar a la semana?*

Una	5,1
Dos	9,8
Tres	12,7
Cuatro	10,6
Cinco	8,2
Seis	3,9
Siete	18,6
Ocho	3,2
Nueve o más	20,3
Utiliza siempre la misma (no es desechable)	3,3
N.S., duda	3,9
N.C.	0,3
<b>(N)</b>	<b>(2.989)</b>
Media (de los que usan mascarillas desechables)	5,50
Desviación típica	2,63
<b>(N)</b>	<b>(2.763)</b>

### **Pregunta 7b**

*Y piense ahora en un día de actividad normal. ¿Cuántas horas suele Ud. usar la mascarilla a lo largo del día, en lugares públicos (calle, comercio, centro de estudios, transporte, etc.) o de trabajo?*

<i>Menos de una hora</i>	<i>5,1</i>
<i>Entre 1 y 2 horas</i>	<i>16,9</i>
<i>Entre 2 y 3 horas</i>	<i>14,0</i>
<i>Entre 3 y 4 horas</i>	<i>11,7</i>
<i>Entre 4 y 5 horas</i>	<i>8,0</i>
<i>Entre 5 y 6 horas</i>	<i>5,8</i>
<i>Entre 6 y 7 horas</i>	<i>3,8</i>
<i>Entre 7 y 8 horas</i>	<i>6,7</i>
<i>Entre 8 y 9 horas</i>	<i>8,7</i>
<i>Entre 9 y 10 horas</i>	<i>7,2</i>
<i>Más de 10 horas</i>	<i>9,9</i>
<i>N.S.</i>	<i>1,8</i>
<i>N.C.</i>	<i>0,3</i>
<i>(N)</i>	<i>(2.989)</i>
<i>Media</i>	<i>4,54</i>
<i>Desviación típica</i>	<i>3,24</i>
<i>(N)</i>	<i>(2.926)</i>

### **Pregunta 8**

*¿Usa Ud. gel hidroalcohólico para desinfectarse las manos?*

<i>Sí</i>	<i>91,6</i>
<i>No</i>	<i>8,3</i>
<i>N.C.</i>	<i>0,0</i>
<i>(N)</i>	<i>(3.008)</i>



### **Pregunta 8a**

*¿Cuándo usa Ud. gel hidroalcohólico para desinfectarse las manos? (RESPUESTA MÚLTIPLE. MARCAR TODAS LAS QUE DIGA LA PERSONA ENTREVISTADA).*

<i>Siempre o casi siempre</i>	35,7
<i>Cada vez que toca algún objeto (botón, pulsador, picaporte, pasamanos...)</i>	34,9
<i>Al llegar al trabajo y durante el trabajo</i>	20,5
<i>Al llegar y al salir de casa, y en casa</i>	31,2
<i>Al entrar o salir de un comercio</i>	57,4
<i>Al entrar o salir de un bar o restaurante</i>	20,4
<i>En cualquier sitio público</i>	3,0
<i>De vez en cuando, en algunas situaciones</i>	2,9
<i>Al subir o bajar de cualquier medio de transporte</i>	3,5
<i>Otras respuestas</i>	0,5
<i>N.C.</i>	0,1
<i>(N)</i>	<i>(2.756)</i>

### **Pregunta 9**

*Aparte de las mascarillas y el gel hidroalcohólico, de las siguientes medidas de protección frente al virus COVID-19, ¿cuáles practica Ud. normalmente?*

	<i>Sí</i>	<i>No</i>	<i>N.C.</i>	<i>(N)</i>
<i>Guarda la distancia de seguridad entre personas recomendada por las autoridades sanitarias</i>	94,4	5,0	0,6	<i>(3.008)</i>
<i>Se lava las manos con frecuencia en casa u otros lugares</i>	95,4	4,6	0,1	<i>(3.008)</i>
<i>Desinfecta productos alimenticios antes de comerlos</i>	44,5	55,2	0,3	<i>(3.008)</i>

### **Pregunta 10**

*¿Utiliza Ud. alguna otra medida de protección frente a la COVID-19?*

<i>Sí</i>	21,0
<i>No</i>	79,0
<i>N.C.</i>	0,1
<i>(N)</i>	<i>(3.008)</i>

### **Pregunta 10a**

*¿Podría decirme cuál o cuáles? (RESPUESTA ESPONTÁNEA). (MÁXIMO TRES RESPUESTAS).*

<i>Ventila en la casa y/o en el trabajo</i>	13,3
<i>Desinfecta la casa/trabajo, objetos, lavadora, utensilios con desinfectante, alcohol, lejía, ozono</i>	24,9
<i>Mantiene la distancia, evita las aglomeraciones, lugares cerrados, vigila los aforos</i>	13,5
<i>Limita las salidas, se confina en casa</i>	4,6
<i>Utiliza guantes, gafas, gorros, pantallas, batas de protección en algunos lugares</i>	15,7
<i>Reduce el contacto social y relaciones sociales</i>	8,9
<i>Desinfecta y limpia zapatos, se descalza, deja fuera los zapatos</i>	17,4
<i>Limpieza y cambio de ropa frecuente, desinfección textil, lavado a alta temperatura</i>	13,9
<i>Aumento de higiene personal y alimentación sana</i>	9,0
<i>Otras respuestas</i>	4,4
<b>(N)</b>	<b>(631)</b>

Los resultados de esta última encuesta, incluso son mejores en referencia al cumplimiento de las medidas y comportamiento para evitar el contagio o propagación del Coronavirus. Más porcentaje de uso de la mascarilla y más número de estas, lo que denota una mayor concienciación sobre la importancia de la efectividad de la misma (*preguntas 7 y 7a*), y más número de horas con ella puesta (*7b*). Y aunque disminuye el uso del gel hidroalcohólico (*8*), en general la puntuación del resto de preguntas son más altas, es decir, mejor cumplimiento de medidas y recomendaciones para la lucha contra la Covid-19.

Tras analizar los resultados de estas encuestas con técnicas directas, se observa, si se realiza un análisis más profundo, una incongruencia en el discurso reproducido por los resultados entre la primera y las dos segundas. Por un lado los valores de la primera nos hablan de desconfianza, de necesidad, de reclamar más dureza en las normas. Por otro lado, la segunda, cuando se le pregunta a la persona por sus hábitos y por su grado de cumplimiento de las medidas de prevención, los valores son muy altos.

El hecho de no cumplir unas normas que son en beneficio de toda la sociedad, sobre todo cuando se trata de la prevención de enfermedades y los riesgos que

conlleva, supone exponerse al juicio social. Es deseable socialmente que todos cumplan las normas. Es por ello que estas preguntas llevan implícito un sesgo de deseabilidad social. Los encuestados, cuando son preguntados por ellas directamente, no contestan con total honestidad, aún habiendo confidencialidad. Se aplica un auto-sesgo de deseabilidad social lo cual impide hacer una estimación real de los resultados de la pregunta.

## 2. Las técnicas indirectas.

De entre las muchas técnicas, a nivel cuantitativo que existen para la recogida de datos y posterior análisis, el presente trabajo se centra en la encuesta, pues es la herramienta que aquí se utiliza. Es preciso explicar antes diferentes métodos, con especial atención a las técnicas indirectas.

La encuesta es un procedimiento de investigación que se basa en interrogar a una muestra de individuos (Cobo Rodríguez, 2013). Ya sea mediante una batería de preguntas, supuestos, valoraciones, etc.

La validez o calidad de los datos es un factor incuestionable para cualquier estudio, esto quiere decir, que necesitamos que la muestra rellene los datos, y que esos datos sean completados de forma sincera, acorde al pensamiento u opinión de la persona encuestada (Cobo Rodríguez, 2013).

En ciertos estudios, puede darse el caso que el tema a analizar sea de características sensibles. Temas los cuales contienen información que no todas las personas confían en informar con total sinceridad. Un problema recurrente es las respuestas en base a la “deseabilidad social”. Este fenómeno es la tendencia de las personas encuestadas a responder en función de lo que está bien valorado socialmente (Cobo Rodríguez, 2013). Se produce incluso cuando se le garantiza el anonimato a la muestra, es una especie de auto-sesgo impuesto ante el temor de ser evaluados.

En estos estudios hay elementos sensibles. Los más destacables, las preguntas en sí. Estas tienen como objeto captar la información y características sensibles de los encuestados. Hay que tener en cuenta las respuestas contestadas de forma no sincera, y las que directamente no han sido contestadas.

A lo largo del pasado siglo surgió la pregunta a este problema: ¿Qué diseño debería tener una encuesta cuyas preguntas busquen información altamente sensible, para que el o la encuestada responda de forma honesta?

El primero en trabajar esta cuestión es Warner (1965) y estableció un planteamiento base con la técnica de respuesta aleatorizada. Más tarde, son destacables las aportaciones de Greenberg et al. (1969), Horvitz et al. (1967), Chaudhuri (1988), Kuk (1990), Mangat (1994) y Chang et al. (2004), que añadieron mejoras al trabajo de Warner. Guo-Liang Tian & Man-Lai Tang (2014) desarrollaron el Modelo Crosswise, que es explicado en profundidad en el presente trabajo por tratarse del modelo con el que se realiza la metodología del análisis de la encuesta acerca del grado de cumplimiento de las medidas contra la transmisión del SARS-CoV-2. El modelo Crosswise cuenta con recientes aportaciones, entre las cuales destaca el “Extended Crosswise Model” (Heck et al., 2018), reciente y simple extensión del Modelo Crosswise que añade análisis de posibles sesgos en muestras no necesariamente muy grandes.

La varianza en todas las técnicas juega un papel fundamental, pues ya sea por método aleatorio o pregunta de forma indirecta, se asume que la varianza aumenta e influye en el cálculo de resultados.

## 2.1. Modelo Warner.

En 1965 Stanley L. Warner decidió diseñar un “modelo de respuesta aleatoria para proporciones múltiples” y establecer un procedimiento diferente de cómo se pregunta en una encuesta directa. Esto hace que los datos también lleguen en otro formato y deban recibir un tratamiento diferente.

Warner demostró que se puede estimar una prevalencia de respuesta sin que la persona encuestada muestre explícitamente su respuesta. Con este modo de completar un cuestionario se evita la ausencia de respuestas o respuestas no sinceras por temor de los encuestados a ser juzgados o que crean que sus respuestas no son confidenciales.

Su método se desarrolla de la siguiente forma:

- Primero se clasifican a las personas en *grupos A y B*.
- $\theta$  es la proporción de personas con características de interés, la característica que interesa analizar en el estudio. Las personas que tienen dicha característica se sitúan en el *grupo A*. El objetivo del método es tener una estimación de  $\theta$  sin preguntar directamente a cada persona si pertenece o no al *grupo A*.
- Se reúne un fajo de cartas en la que una parte de ellas  $p$  se marca con una con la letra *A (grupo A)*, y la parte restante  $1-p$ , con el resto de letras del abecedario (*grupo B*).
- Se selecciona una muestra aleatoria simple o estratificada con reemplazo de un tamaño  $n$ , perteneciente a una población de tamaño ( $N$ ).
- Se muestra el conjunto de cartas a cada individuo de la muestra para que vean que están marcadas con las diferentes letras del abecedario.
- Se le da la instrucción de que responda *SÍ* o *NO* a la pregunta que se le haga, haciendo hincapié en que debe de estar muy atento a la pregunta.
- Si al coger una carta, está marcada con la letra *A*, entonces tiene que responder a la pregunta formulada de forma afirmativa, esto es: *¿Tiene la característica sensible?*. Si al coger una carta, está marcada con alguna de las letras restantes del abecedario, entonces tiene que responder a la pregunta formulada de forma negativa, sería: *¿No tiene la característica sensible?*.
- Hay que insistir en que la sinceridad de la respuesta es crucial. Aún independientemente de cual letra haya salido en la carta que aleatoriamente se extrae.
- Después de que se entreviste a cada individuo hay que reemplazar la carta al fajo. Es decir, hay que volver a introducirla para que pueda ser extraída de nuevo en cada caso.
- El procedimiento se aplica por igual a toda la muestra  $n$ .
- Con las  $n$  respuestas de cada categoría *sí* y *no* de cada grupo, se hacen estimaciones para ver la prevalencia de la característica sensible del tema de estudio.

Warner presenta el método valiéndose de una baraja de cartas como herramienta de aleatorización, pero puede valer cualquiera, desde un dado hasta software informático diseñado para ello. Lo importante es saber qué probabilidad se le otorga al *grupo A*. Esta puede ser elegida a gusto del encuestador, pero no debe ser  $p = 1/2$ , ni tampoco  $p = 1$ . En el primer caso, la probabilidad sería totalmente aleatoria, con lo que no se podría realizar ninguna estimación posterior. En el segundo caso, el encuestado podría darse cuenta de que directamente se le está preguntando si pertenece o no al *grupo A*. Es decir, una pregunta, formulada afirmativamente, de forma directa, con los sesgos de deseabilidad social que ello conlleva. Esto sería una técnica directa en lugar de respuesta aleatorizada. Darle a *A* una probabilidad de  $3/4$  es bastante usual. Esto es así porque se necesita un tamaño muestral  $n$  bastante grande para hacer buenas estimaciones, cada respuesta no da tanta información como cuando se pregunta de forma directa.

Así se define la variable de interés:

- $y_i = 1$  si la persona tiene la característica *A* (pertenece al *grupo A*).
- $y_i = 0$  si la persona tiene la característica  $A^c$  (que no tiene la característica *A* y pertenece al *grupo B*).

Lo que se pretende es una estimación correcta de  $\theta = Y/N$ , ( $Y = \sum_1^N y_i$ ) obteniendo *A*.

Se definirá una variable que toma valores  $I_i = 1$  si hay coincidencia entre la carta extraída por el encuestado y que la persona tenga la característica *A*. Lo opuesto sería  $I_i = 0$  y la característica  $A^c$ .

En el modelo de Warner  $E_R$ ,  $V_R$  representan la esperanza y la varianza, bajo el procedimiento aleatorio, respectivamente. En la técnica de Respuesta Aleatorizada se deduce que:

$$E_R(I_i) = py_i + (1 - p)(1 - y_i) = (1 - p) + (2p - 1)y_i$$

lo que lleva a:

$$r_i = \frac{I_i - (1-p)}{2p-1}$$

junto con:

$$E_R(r_i) = y_i$$

Además, puesto que  $I_i^2 = I_i$  y  $y_i^2 = y_i$ ,

$$V_R(I_i) = E_R(I_i) (1 - E_R(I_i)) = p(1 - p),$$

por tanto,

$$V_i = V_R(r_i) = \frac{V_R(I_i)}{(2p-1)^2} = \frac{p(1-p)}{(2p-1)^2} = \phi_w$$

Siendo  $r_i$  variables independientes en todo  $i$ , con  $\mu = (1, \dots, i, \dots, N)$ .

Warner sólo contemplaba el muestreo aleatorio simple con reemplazamiento en su teoría. A partir de cada persona seleccionada y sus respuestas aleatorias, se obtiene un estimador insesgado para  $\theta$ .

Por lo que,

$$r_K = \frac{I_K - (1-p)}{2p-1}, K = 1, \dots, n$$

tiene la esperanza

$$E_R(r_K) = y_K$$

y

$$\bar{r} = \frac{1}{n} \sum_{K=1}^n r_k$$

tiene la esperanza

$$E_R(\bar{r}) = \frac{1}{n} \sum_{K=1}^n y_k = \bar{y}$$

También,



$$V_R(\bar{r}) = \frac{\phi_w}{n}$$

$E_p, V_p$  son los operadores genéricos de la esperanza y la varianza, respectivamente en cualquier diseño de muestreo  $p$ . Entonces:

$$E = E_p E_R = E_R E_p$$

$$V = E_p V_R + V_p E_R = E_R V_p + V_R E_p,$$

se obtiene,

$$E(\bar{r}) = E_p(\bar{y}) = \frac{1}{N} \sum_1^N y_i = \frac{Y}{N} = \bar{Y} = \theta$$

en este caso,

$$V(\bar{r}) = E_p\left(\frac{\phi_w}{n}\right) + V_p(\bar{y}) = \frac{\phi_w}{n} + \frac{\sigma^2}{n}$$

Al escribir:

$$\sigma^2 = \frac{1}{N} \sum_1^N (y_i - \bar{Y})^2$$

y como  $y_i^2 = y_i$ , Warner consiguió

$$V(r) = \frac{1}{n} [\Phi_w + \theta(1 - \theta)]$$

Escribiendo

$$S_r^2 = \frac{1}{n-1} \sum_{K=1}^n (r_K - r)^2,$$

resulta:

$$v = \frac{1}{n} S_r^2$$

que es un estimador sin sesgo para  $V(\bar{r})$ . Warner observa que  $\sum_{i=1}^n I_i$  sigue una distribución binomial  $b(n, \lambda)$  con

$$\lambda = p\theta + (1 - p)(1 - \theta) = (1 - p) + (2p - 1)\theta$$

$\lambda$  es la probabilidad de que una persona, en cada extracción, en el muestreo aleatorio simple, pueda responder “sí” con honestidad.

Teniendo en cuenta que

$$n_1 = \sum_{i=1}^n I_i$$

es el número total de observaciones de respuestas afirmativas en  $n$  extracciones, con el muestreo ya mencionado, y

$$\hat{\lambda} = \frac{n_1}{n},$$

Warner (1965) calculó el estimador sin sesgo de  $v(\bar{r})$ :

$$v' = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2p-1)^2}$$

Se puede comprobar fácilmente que  $v = v'$ .

Aplicación en un diseño muestral general.

Se recoge una muestra de personas  $s$ , de la población  $U$ , con una probabilidad  $p(s)$  de acuerdo a una probabilidad de diseño  $p$ . En esta  $p$  la probabilidad de inclusión  $i$  es  $\pi_i = \sum_{s \ni i} p(s)$ ,  $i \in U$  y para un par distinto de individuos  $i, j (i \neq j)$  es  $\pi_{ij} = \sum_{s \ni i, j} p(s)$ . Se restringen los diseños para que  $\pi_i > 0 \forall i, j \in U, i \neq j$ .

Por lo tanto, se estima  $\theta$  por

$$e = \frac{1}{N} \sum_{i \in S} \frac{r_i}{\pi_i} = \frac{1}{N} \left[ \sum_{i \in S} \frac{I_i - (1-p)}{\pi_i} \right] = \frac{1}{N} \frac{1}{2p-1} \left[ \sum_{i \in S} \frac{I_i}{\pi_i} - (1-p) \sum_{i \in S} \frac{1}{\pi_i} \right],$$

sabiendo que,

$$V(e) = V_p(t) + \sum \frac{V_i}{\pi_i}$$

siendo también la varianza de Horvitz y Thompson:

$$V_p(t) = \sum_{i < j}^N \sum_j^N \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 (\pi_i \pi_j - \pi_{ij}) + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i$$

entonces:

$$\beta_i = 1 + \frac{1}{\pi_i} \sum_{i \neq j} \pi_{ij} - \sum_1^N \pi_i$$

se obtiene:

$$N^2 V(e) = \sum_{i < j}^N \sum_j^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \phi_p \sum_1^N \frac{1}{\pi_i}$$

siendo:

$$V_i = V_R(r_i) = \frac{p(1-p)}{(2p-1)^2} = \phi_p$$

y:

$$N^2 \hat{V}(e) = \sum_{i < j} \sum_i (\pi_i \pi_j - \pi_{ij}) \frac{I_{sij}}{\pi_{ij}} \left( \frac{r_i}{\pi_i} - \frac{r_j}{\pi_j} \right)^2 + \sum \frac{r_i^2}{\pi_i} \beta_i \frac{I_{si}}{\pi_i} + \phi_p \sum \frac{I_{si}}{\pi_i}$$

siendo:

$$I_{sij} = I_{si} I_{sj} = \begin{cases} 1 & \text{si } i, j \text{ ambos están en } s \\ 0 & \text{en otro caso} \end{cases}$$

## 2.1.2. Ampliaciones y mejoras destacables del Modelo Warner.

Modelo de Respuesta no relacionada.

Este modelo también se basa en dos preguntas, pero ahora se trata de una que se interesa por el atributo sensible, y otra que se interesa por algún otro aspecto, no sensible, y no relacionado con el primero. Sobre esta última, existe una probabilidad conocida de que sea afirmativa, la cual es  $\alpha$  (Greenberg et al., 1969).

La probabilidad de la pregunta sensible es  $p$  de nuevo, y la de la inocua es  $1 - p$ .

Se deduce que, en general:

$$I_i = \begin{cases} y_i & \text{con probabilidad } p \\ w_i & \text{con probabilidad } 1 - p \end{cases}$$

$$E_R(I_i) = py_i + (1 - p)\alpha,$$

siendo,

$$r_i = \frac{I_i - (1 - p)\alpha}{p}$$

y

$$E_R(r_i) = y_i$$

entonces

$$V_R(I_i) = E_R(I_i^2) - E_R^2(I_i) = p(1 - p)(y_i - \alpha)^2$$

$$V_R(r_i) = \frac{V_R(I_i)}{p^2} = \frac{p(1 - p)(y_i - \alpha)^2}{p^2}$$

Se estima  $\theta$  por:

$$e = \frac{1}{N} \sum_{i \in S} \frac{r_i}{\pi_i} = \frac{1}{N} \sum_{i \in S} \frac{I_i - (1-p)\alpha}{p} = \frac{1}{N} \frac{1}{p} \left[ \sum_{i \in S} \frac{I_i}{\pi_i} - (1-p)\alpha \sum_{i \in S} \frac{1}{\pi_i} \right]$$

y

$$\begin{aligned} N^2 V(e) &= \sum_{i < j}^N \sum_j^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \sum_1^N \frac{p(1-p)(y_i - \alpha)^2}{p^2 \pi_i} \\ &= \sum_{i < j}^N \sum_j^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \frac{p(1-p)}{p^2} \sum_1^N \frac{(y_i - \alpha)^2}{\pi_i} \\ &= \sum_{i < j}^N \sum_j^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \frac{1-p}{p} \sum_1^N \frac{(y_i - \alpha)^2}{\pi_i} \end{aligned}$$

El contraste entre los Modelos de Warner y el Modelo de Greenberg ha sido realizado por Moors (1971) en poblaciones infinitas, siendo el segundo más eficiente que el primero.

#### Modelo C

En este modelo se mejora la precisión de los estimadores introduciendo información auxiliar correlacionada con la variable de interés. Aquí también hay una variable sensible ( $y$ ) y otra inocua ( $W$ ). La diferencia con el modelo anterior, es que ambas variables están correlacionadas, pero no es sensible, por lo que no compromete la confidencialidad.

Trabajando sobre muestreos finitos, el procedimiento es como en el Modelo Greenberg, aunque se aprovecha la información que se obtiene de la correlación entre las variables sensibles e inocuas.

Se encuentra:

$$E_R(I_i) = py_i + (1 - p)\alpha,$$

siendo:

$$r_i = \frac{I_i - (1 - p)\alpha}{p}$$

y

$$E_R(r_i) = y_i$$

Se estima  $\theta$ :

$$e = \frac{1}{N} \sum_{i \in S} \frac{r_i}{\pi_i} = \frac{1}{N} \sum_{i \in S} \frac{\frac{I_i - (1 - p)\alpha}{p}}{\pi_i} = \frac{1}{N} \frac{1}{p} \left[ \sum_{i \in S} \frac{I_i}{\pi_i} - (1 - p)\alpha \sum_{i \in S} \frac{1}{\pi_i} \right]$$

La varianza del estimador se minimiza si aumenta la correlación entre  $W$  y  $y$ .

Modelo de respuesta forzada.

Se trata de una alternativa al procedimiento que Warner que otorga mayor anonimato al encuestado, y que no recurre a pregunta complementaria.

Cada individuo de la muestra selecciona al azar una de estas tres propuestas:

- 1 - Sensible.
- 2 - Instrucción que dice "sí".
- 3 - Instrucción que dice "no".

Sus probabilidades son  $p_1$ ,  $p_2$ , y  $p_3$  y  $p_1 + p_2 + p_3 = 1$ .

Pero este modelo también tiene estas restricciones:

$$\begin{aligned}
1 > p = p_1 > \frac{1}{2}, \\
0 < 1 - p_1 - p_2 \leq p_2 < 1 - p_1 \\
&\Leftrightarrow \\
1 - p_1 - 2p_2 \leq 0, \\
p_2 < 1 - p_1
\end{aligned}$$

Se tiene

$$I_i = \begin{cases} y_i & \text{con probabilidad } p_1, \\ 1 & \text{con probabilidad } p_2, \\ 0 & \text{con probabilidad } 1 - p_1 - p_2, \end{cases}$$

donde

$$E_R(I_i) = p_1 y_i + p_2$$

siendo

$$r_i = \frac{I_i - p_2}{p_1}$$

y

$$E_R(r_i) = y_i$$

Por tanto

$$\begin{aligned}
V_R(I_i) &= E_R(I_i^2) - E_R^2(I_i) = E_R(I_i) - E_R^2(I_i) = p_1 y_i + p_2 - (p_1 y_i + p_2)^2 \\
&= p_1 y_i + p_2 - (p_1^2 y_i^2 + p_2^2 + 2p_1 y_i p_2) \\
&= p_1 y_i + p_2 - p_1^2 y_i^2 - p_2^2 - 2p_1 y_i p_2 \\
&= p_1 y_i (1 - p_1 - 2p_2) + p_2 (1 - p_2) \\
V_R(r_i) &= \frac{V_R(I_i)}{p_1^2} = \frac{p_1 y_i (1 - p_1 - 2p_2) + p_2 (1 - p_2)}{p_1^2} \\
&= \frac{y_i (1 - p_1 - 2p_2)}{p_1} + \frac{p_2 (1 - p_2)}{p_1^2}
\end{aligned}$$

Entonces, estimamos  $\theta$  por:

$$e = \frac{1}{N} \sum_{i \in S} \frac{r_i}{\pi_i} = \frac{1}{N} \sum_{i \in S} \frac{I_i - p_2}{p_1 \pi_i} = \frac{1}{N p_1} \left[ \sum_{i \in S} \frac{I_i}{\pi_i} - p_2 \sum_{i \in S} \frac{1}{\pi_i} \right]$$

y

$$\begin{aligned} N^2 V(e) &= \sum_{i < j}^N \sum_j^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i \\ &+ \sum_1^N \frac{\frac{y_i(1-p_1-2p_2)}{p_1} + \frac{p_2(1-p_2)}{p_1^2}}{\pi_i} \\ &= \sum_{i < j}^N \sum_j^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \frac{1-p_1-2p_2}{p_1} \sum_1^N \frac{y_i}{\pi_i} \\ &+ \frac{p_2(1-p_2)}{p_1^2} \sum_1^N \frac{1}{\pi_i} \end{aligned}$$

Modelo de Devore

Jay L. Devore (1977) hace una propuesta en la misma línea de Greenberg. Pero, esta vez, pertenecer al grupo inocuo ( $W$ ) se establece con probabilidad 1. Se tiene:

$$I_i = \begin{cases} y_i & \text{con probabilidad } p \\ 1 & \text{con probabilidad } 1-p \end{cases}$$

donde:

$$E_R(I_i) = py_i + (1-p)$$

siendo:

$$r_i = \frac{I_i - (1-p)}{p}$$



y:

$$E_R(r_i) = y_i$$

Entonces:

$$V_R(I_i) = E_R(I_i^2) - E_R^2(I_i) = p(1-p)(1-y_i)$$

$$V_R(r_i) = \frac{V_R(I_i)}{p^2} = \frac{p(1-p)(1-y_i)}{p^2}$$

Se estima  $\theta$ :

$$e = \frac{1}{N} \sum_{i \in S} \frac{r_i}{\pi_i} = \frac{1}{N} \sum_{i \in S} \frac{I_i - (1-p)}{p \pi_i} = \frac{1}{N} \frac{1}{p} \left[ \sum_{i \in S} \frac{I_i}{\pi_i} - (1-p) \sum_{i \in S} \frac{1}{\pi_i} \right]$$

y

$$\begin{aligned} N^2 V(e) &= \sum_{i < j}^N \sum_{j}^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \sum_1^N \frac{p(1-p)(1-y_i)}{p^2 \pi_i} \\ &= \sum_{i < j}^N \sum_{j}^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \frac{p(1-p)}{p^2} \sum_1^N \frac{1-y_i}{\pi_i} \\ &= \sum_{i < j}^N \sum_{j}^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum_1^N \frac{y_i^2}{\pi_i} \beta_i + \frac{1-p}{p} \sum_1^N \frac{1-y_i}{\pi_i} \end{aligned}$$

## 2.2. Modelo Crosswise.

Vistos varios de los modelos más importantes de técnicas de respuesta aleatoria, es preciso ahora centrarse en el Modelo Crosswise, pues va a ser el que se emplee para llevar a cabo el trabajo de campo y análisis del presente trabajo de investigación.

El Modelo Crosswise es una de las más recientes extensiones del Modelo Warner, como método de reducción del sesgo en cuestiones que se ven afectadas por la deseabilidad social. Este modelo no tiene una respuesta autoprotectora para el encuestado. Ni tampoco algún dispositivo aleatorio como dado o baraja de cartas, por eso, el modelo Crosswise es una versión “no aleatorizada” del Modelo Warner (Tian & Tang, 2014).

Este modelo trabaja con dos preguntas en bloque. Una pregunta sobre el atributo sensible que esté siendo estudiado en el trabajo de investigación. Otra pregunta sobre algún atributo no sensible, el cual el investigador/a, previamente conozca una estimación general para la población estudiada.

Se pregunta entonces, en bloques, cada bloque contiene dos preguntas, la sensible y la no sensible. Las posibles respuestas en base a esta metodología de pregunta, teniendo en cuenta que se desea que la persona encuestada mantenga su sinceridad y máxima confidencialidad, son:

- El encuestado responde SÍ o NO a ambas preguntas, simultáneamente. Es decir, responde SÍ a la pregunta sensible y a la no sensible. O bien, responde NO a la pregunta sensible y a la no sensible. Esta forma de presentar el cuestionario otorga mucha más tranquilidad al encuestado y sensación de confidencialidad. Lo que permite unas estimaciones mucho más precisas.
- El encuestado responde SÍ a una de las preguntas y NO a la otra pregunta. No importa cuál es SÍ y cuál es NO, si la primera o la segunda. Solo tiene que indicar que tienen una respuesta diferente.

Se puede ver más visualmente así:

	<b>Responder “SÍ” a la pregunta con atributo no sensible</b>	<b>Responder “NO” a la pregunta con atributo no sensible</b>
<b>Responder “SÍ” a la pregunta con atributo sensible</b>	<i>Posibilidad 1 → A</i>	<i>Posibilidad 2 → B</i>
<b>Responder “NO” a la pregunta con atributo sensible</b>	<i>Posibilidad 2 → B</i>	<i>Posibilidad 1 → A</i>

Tabla: Modelo de respuesta para Modelo Crosswise (Giménez Luque, 2020).

Esta primera definición de caracteres ayuda a comprender las fórmulas que definen probabilidad y estimaciones de más adelante:

$\pi$ = Proporción de población con características sensibles.

$P$ = Proporción de población con características no-sensibles.

$n$ = Número de encuestados.

$A$ = Se llamará *Respuesta A* a la *Posibilidad 1* especificada en la tabla de más arriba ( ambas SÍ o ambas NO).

$B$ = Se llamará *Respuesta B* a la *Posibilidad 2* especificada en la tabla de más arriba ( una SÍ, otra NO).

$\lambda$ = Proporción esperada de respuestas A en la población.

$\hat{\pi}$ = Proporción estimada de población con características sensibles.

$\hat{\lambda}$ = Proporción observada de respuestas A en la población ( $A/n$ ).

El siguiente árbol de probabilidad ayuda a ver la probabilidad con la que un encuestado puede contestar  $A$  o  $B$ , donde 1 corresponde a encuestados con características sensibles; 3 y 5 a encuestados con características no sensibles; 2 a encuestados sin características sensibles; 4 y 6 a encuestados sin características no sensibles:

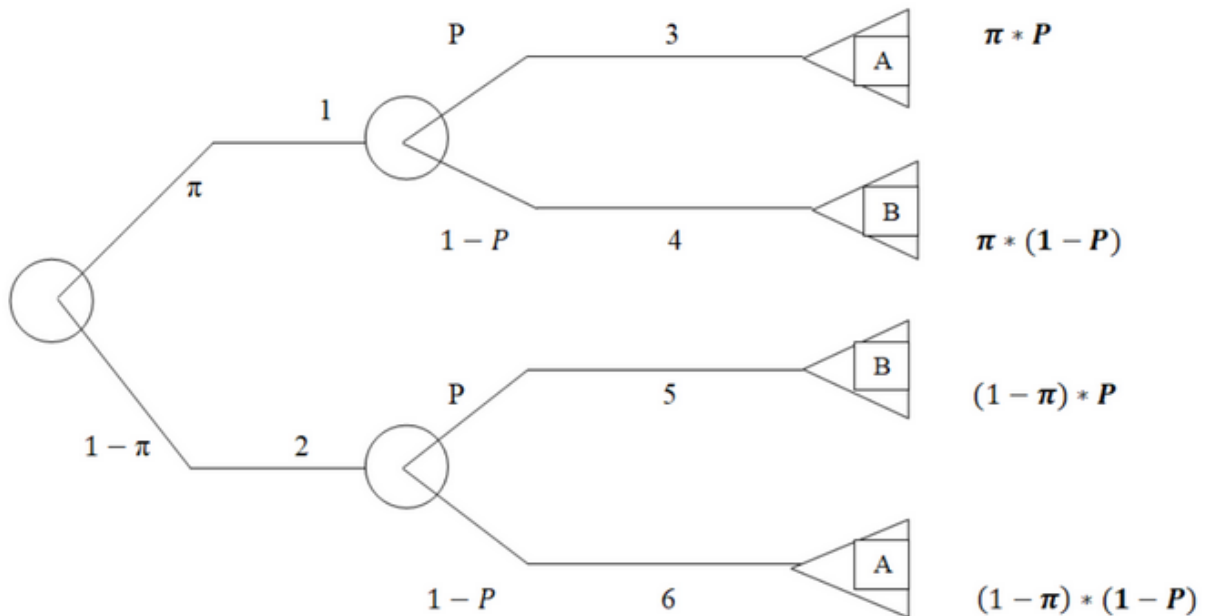


Figura: Árbol probabilidad Modelo Crosswise (Giménez Luque, 2020).

Luego, la estimación de la proporción de la muestra que posee la característica sensible, es:

$$\hat{\pi} = \frac{(\hat{\lambda} + P - 1)}{2 * P - 1}$$

Y la estimación de la varianza es:

$$\widehat{var}(\hat{\pi}) = \frac{\hat{\lambda} * (1 - \hat{\lambda})}{(n - 1) * (2 * P - 1)^2} = \frac{\hat{\pi} * (1 - \hat{\pi})}{n - 1} + \frac{P * (1 - P)}{(n - 1) * (2 * P - 1)^2}$$

Estos estimadores pueden ser calculados si se conoce la estimación de la población para la pregunta no sensible y con la cantidad de respuestas  $A$  y  $B$ .

### 2.2.1. Variación del Modelo Crosswise: el modelo triangular.

La diferencia respecto al modelo crosswise es en la forma en la que el encuestado responde. En este caso, si la respuesta a ambas preguntas del bloque es *NO* el encuestado responderá *A*, para el resto de posibilidades contestará *B*. Es decir si las dos preguntas son *SÍ*, o una de ellas, independientemente de cual sea, es *SÍ* y la otra *NO*.

	SÍ a la pregunta sensible	NO a la pregunta sensible
SÍ a la pregunta no sensible	<i>B</i>	<i>B</i>
NO a la pregunta no sensible	<i>B</i>	<i>A</i>

Tabla: Modelo de respuesta para modelo triangular (Giménez Luque, 2020).

Se especifica ahora también una aclaración de los elementos que van a aparecer en las fórmulas que expliquen el cálculo de estimaciones en este modelo:

$\pi$ = Proporción de población con características sensibles

$P$ = Proporción de población con características no-sensibles

$n$ = Número de encuestados

$A$ = Respuesta A (ambas NO).

$B$ = Respuesta B (al menos una SÍ).

$\lambda$ = Proporción esperada de respuestas A en la población.

$\hat{\pi}$ = Proporción estimada de población con características sensibles.

$\hat{\lambda}$ = Proporción observada de respuestas A en la población ( $A/n$ ).

Como antes, el siguiente árbol de probabilidad ayuda a ver la probabilidad con la que un encuestado puede contestar *A* o *B*, donde 1 corresponde a encuestados con

características sensibles; 3 y 5 a encuestados con características no sensibles; 2 a encuestados sin características sensibles; 4 y 6 a encuestados sin características no sensibles:

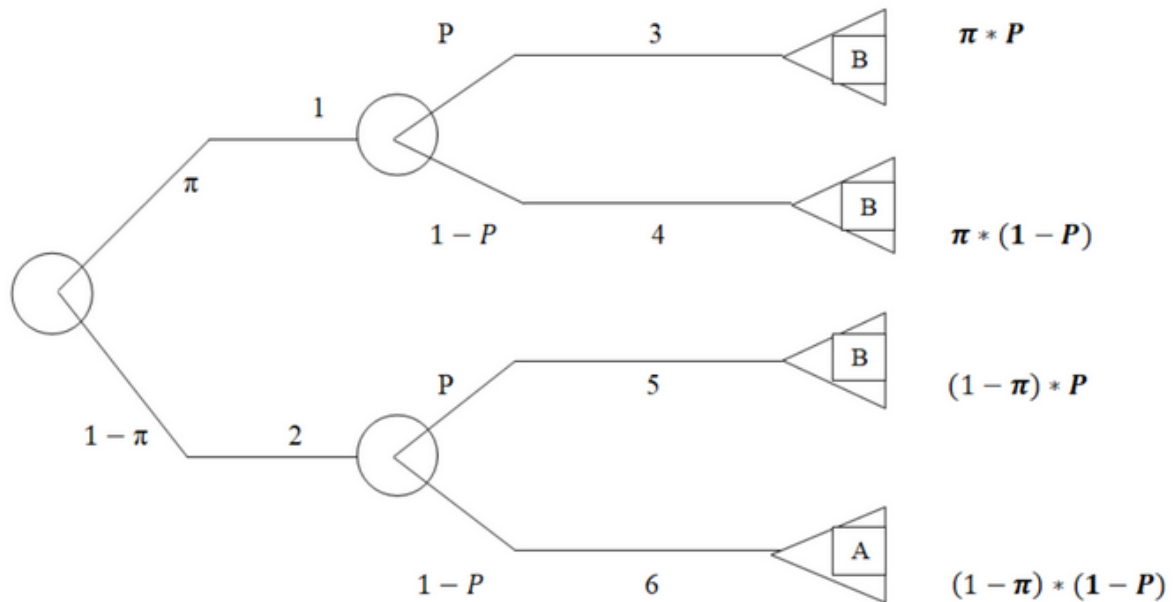


Figura: Árbol probabilidad modelo triangular (Giménez Luque, 2020).

La estimación de la proporción de individuos que poseen la característica sensible se expresa así:

$$\hat{\pi} = 1 - \frac{\hat{\lambda}}{1 - P}$$

La estimación de la varianza:

$$\widehat{var}(\hat{\pi}) = \frac{\hat{\lambda} * (1 - \hat{\lambda})}{(n) * (1 - P)^2}$$

Con este método también se puede calcular la estimación poblacional del atributo sensible, si previamente ya sabemos la prevalencia del atributo no sensible en la población. Como ya se ha dicho, es un modelo de respuesta aleatoria, entonces:

$P((Y_i^*) = 1)$  representa la probabilidad de que la respuesta sea Sí. Lo que en este modelo sería *Respuesta B*.

$P((Y_i) = 1)$  representa la probabilidad de una respuesta afirmativa, pero, en este caso representada por  $\pi$ .

Por tanto, la probabilidad de la *Respuesta B* sería:

$$P(Y_i^* = 1) = \pi * p + \pi(1 - p) + (1 - \pi) * p$$

$$P(Y_i^* = 1) = \pi * p + \pi - \pi * p + p - p * \pi$$

$$P(Y_i^* = 1) = \pi * (1 - p) + p$$

$$P(Y_i^* = 1) = c + d * \pi$$

La ecuación es similar al modelo de respuesta aleatoria, en el que  $c$  es igual a  $p$  y  $d$  es igual a  $(1 - p)$ .

## 2.3. Regresión logística para el Modelo Crosswise.

Van den Hout et al. (2007) distingue entre dos métodos para llevar a cabo el análisis de las observaciones obtenidas con el Modelo Crosswise. Método para respuestas aleatorias univariantes y método para respuestas aleatorias multivariantes.

### 2.3.1. Regresión logística en respuestas aleatorias univariantes.

Primero se definen las variables adicionales que se utilizan para las fórmulas de análisis de este modelo:

$Y$  = variable dependiente observada ( $Y = 1$  vs  $Y = 0$ )

$Y^*$  = variable binaria verdadera (sensible)

$c$  = Probabilidad de una respuesta afirmativa forzada

$d$  = Probabilidad de una respuesta verdadera

$x$  = Predictor

$\beta_0$  = Constante

$\beta_1$  = Coeficiente de regresión del predictor

$p$  = Probabilidad de aleatorización

$i$  = índice referente al encuestado  $i$

$k$  = índice referente a la pregunta número  $k$

Hay que tratar de establecer correctamente una relación entre la variable sensible, es decir, la de estudio y los diferentes predictores arriba definidos. Los datos que se recopilan durante el trabajo de campo son de carácter binario (sólo hay dos caracteres posibles en las respuestas), es por ello hay que realizar una transformación logística, por lo que se usa la siguiente función logística:



$$\log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \beta_0 + \beta_1 * x_i$$

Esta función es llamada  $\text{Log}(P(Y_i))$ , pues es una transformación logarítmica de las probabilidades, y representa probabilidad, tanto de éxito como de fracaso. Aunque esta función todavía no contempla ninguna propiedad de respuesta aleatoria. Hay que advertir que se observa contaminación de ruido en la variable subyacente por el propio método aleatorio. Esto sucede en todos los métodos de respuesta aleatoria, y se representa con la siguiente función:

$$P(Y_i^* = 1) = c + d * P(Y_i = 1).$$

Para solucionar esto hay que crear la siguiente función de enlace:

$$\log(g(\text{probabilidades}(Y_i))) = \log \frac{P(Y_i^* = 1) - c}{d - P(Y_i^* = 1) + c}$$

Cuyos pasos intermedios son:

$$P(Y_i = 1) = \frac{P(Y_i^* = 1) - c}{d}$$

$$P(Y_i = 0) = 1 - P(Y_i = 1) = \frac{d - P(Y_i^* = 1) + c}{d}$$

$$\log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \log \frac{\frac{P(Y_i^* = 1) - c}{d}}{\frac{d - P(Y_i^* = 1) + c}{d}} = \log \frac{P(Y_i^* = 1) - c}{d - P(Y_i^* = 1) + c}$$

$c$  y  $d$  están definidas por el método de respuesta aleatoria, y en cada método se le otorgan unos valores. Aquí en el modelo Crosswise sería:

$$c = 1 - p$$

$$d = 2p - 1$$

Se muestra ahora una tabla con la parametrización de las diferentes funciones de enlace para los datos binarios distribuidos, realizada por Bernoulli.

	<b>Función</b>	<b>Logit</b>	<b>Probit</b>
<b>Distribución</b>	$\pi(x_j^t \beta)$	Logistic	Cum. Normal
<b>Valor Esperado</b>	$\mu_i$	$c + \frac{d \exp(\eta_i)}{1 + \exp(\eta_i)}$	$c + d \Phi(\eta_i)$
<b>Link</b>	$g(\mu_i)$	$\log\left(\frac{\mu_i - c}{c + d - \mu_i}\right)$	$\Phi^{-1}\left(\frac{\mu_i - c}{d}\right)$
<b>Link inverso</b>	$g^{-1}(\eta_i)$	$c + \frac{d \exp(\eta_i)}{1 + \exp(\eta_i)}$	$c + d \Phi(\eta_i)$
<b>Derivada</b>	$\frac{d g^{-1}(\eta_i)}{d \eta_i}$	$\frac{d \exp(\eta_i)}{(1 + \exp(\eta_i))^2}$	$d \phi(\eta_i)$
	<b>Compl. Log-log</b>	<b>Cauchit</b>	
<b>Distribución</b>	Gumbel	Cauchy	
<b>Valor Esperado</b>	$c + d(1 - \exp(1 - \exp(\eta_i)))$	$c + d(\arctan(\eta)/\pi + \frac{1}{2})$	
<b>Link</b>	$\log(-\log(\frac{c+d-\mu_i}{d}))$	$\tan(\pi(\frac{1}{d}(\mu_i - c)))$	
<b>Link inverso</b>	$c + d(1 - \exp(1 - \exp(\eta_i)))$	$c + d(\arctan(\eta)/\pi + \frac{1}{2})$	
<b>Derivada</b>	$d \exp(\eta_i) \exp(-\exp(\eta_i))$	$d / (\pi(1 + \eta^2))$	

Nota:  $\eta_i = \beta^t x_i$ ,  $g(\mu_i)$ ,  $\arctan(.) = \tan^{-1}(.)$

Tabla: Parametrización de las diferentes funciones de enlace para los datos RR binarios distribuidos de Bernoulli (Giménez Luque, 2020).

### 2.3.2. Estimadores para el Modelo Crosswise.

Las ecuaciones son las mismas para los modelos en general de respuestas aleatorias como lo que no lo son, cuando se trata de modelos mixtos lineales generalizados. Vienen a ser:

$$E(y_i | x_i) = \mu_i = c + dP(Y = 1 | x_i) = c + d g^{-1}(\eta_i) = c + d g^{-1}(x_i^t \beta).$$

La diferencia radica en que la probabilidad de éxito incluye los parámetros para respuestas aleatorias  $c$  y  $d$ . Siendo  $\pi(x^t \beta)$  los predictores lineales de la verdadera

probabilidad de éxito, que se transforman linealmente, lo cual no modifica las ecuaciones arriba formuladas. Aunque la diferencia es que los componentes de las ecuaciones también contienen parámetros de diseño.

Por tanto, las funciones de enlace para datos binarios distribuidos, no cambian los resultados en las ecuaciones de estimación, pues son solamente el puente entre la respuesta final esperada y el predictor lineal.

Entonces, la función de log-verosimilitud se establece como:

$$\log L(\beta; y) = \sum_{i=1}^n y_i \log(c + d\pi(x_i^t \beta)) + (1 - y_i) \log(1 - (c + d\pi(x_i^t \beta)))$$

Se comprueba que es la misma función log-verosimilitud que la del modelo mixto lineal generalizado para datos binarios, solo cambia la probabilidad de éxito. Por ello, la función de puntuación y matriz de información para modelos de respuestas aleatorias, queda como la del modelo mixto lineal generalizado, sería:

$$S(\beta; y) = \frac{\partial \log L(\beta; y)}{\partial \beta} = \sum_{i=1}^n \frac{x_i y_i}{c + d\pi(x_i^t \beta)} \frac{d\pi(\eta_i)}{\partial \eta_i} - \frac{x_i (1 - y_i)}{1 - (c + d\pi(x_i^t \beta))} \frac{d\pi(\eta_i)}{\partial \eta_i} = \sum_{i=1}^n x_i * \frac{d\pi(\eta_i)}{\partial \eta_i} \left[ \frac{y_i - (c + d\pi(x_i^t \beta))}{(c + d\pi(x_i^t \beta))(1 - (c + d\pi(x_i^t \beta)))} \right] = X^t D \sum_{i=1}^n (y - \mu)$$

Características de la función:

- La matriz de diseño está dada por:  $X^t = (X_1, \dots, X_n)$ .
- Las derivadas de primer orden están dadas por:  
 $D = \text{Diag}(d\pi(\eta_1)/\partial \eta, \dots, d\pi(\eta_n)/\partial \eta)$
- La matriz de covarianza por:  $\Sigma = \text{Diag}(\sigma_1, \dots, \sigma_n)$ , donde  
 $\sigma_i = (c + d\pi(\eta_i))(1 - (c + d\pi(\eta_i)))$ .

La varianza asintótica se define:

$$I(\beta \parallel y) = E\left(\sum_{i=1}^n S(\beta; y_i)S(\beta; y_i)^t\right) = \frac{\sum_{i=1}^n X_i X_i^t \left(\frac{d\pi(\eta_i)}{d\eta_i}\right)^2}{\sigma_i} = X^t (D \sum^{-1} D^t) X^t,$$

Que es muy parecida a la matriz de información para el modelo lineal generalizado.

La covarianza asintótica del estimador máximo verosímil  $\hat{\beta}$  es:

$$var(\hat{\beta}) = (X^t (\hat{D} \hat{\Sigma}^{-1} \hat{D}^t) X)^{-1}$$

La peculiaridad aquí es que el término de varianza es diferente, pues se basa en probabilidades de éxito. También que los elementos se multiplican por el parámetro de diseño para respuestas aleatorias  $d$ .

Aclarar que en la última fórmula:

- La matriz  $\hat{D}$  se evalúa en  $\beta = \hat{\beta}$ , como matriz de derivados de primer orden.
- La matriz  $\hat{\Sigma}$  también se evalúa en  $\beta = \hat{\beta}$ , como matriz de covarianza.

### 2.3.3. Regresión logística para respuestas aleatorias multivariantes.

Al haber más variables, habrá más riqueza de información y datos recopilados. Lo cual permitirá unas mejores estimaciones en los análisis realizados, para saber si los atributos sensibles se dan en mayor o menor medida en los sujetos de la muestra. Eso sí, todo depende de la variable sensible latente. Como resumen, se puede decir que, al usar el modelo lineal generalizado, con múltiples variables dependientes, se usa este modelo:

$$\log(g(P(Y_i k))) = \beta_0 k + \beta_1 k x_i$$

## 2.4. Ejemplo de aplicación del Método Crosswise en un estudio sobre el Coronavirus.

Mieth et al. (2020) llevaron a cabo un estudio aplicando la Técnica Crosswise. Precisamente, el objetivo de su investigación era saber si el sesgo de deseabilidad social interferiría en las estimaciones, a la hora de preguntar de forma directa, por el cumplimiento de una de las medidas de prevención de transmisión del coronavirus: el lavado de manos.

### Metodología empleada.

La encuesta estuvo activa desde el 26 al 30 de marzo online en un enlace alojado en el servidor de la Universidad de Düsseldorf. Estaba en alemán y el único requisito era ser capaz de responder la encuesta, de entender el texto. Se obtuvieron aproximadamente 1.300 encuestas válidas.

Hubo dos encuestas, la encuesta con preguntas directas y la encuesta con preguntas indirectas.

- **La encuesta directa:** preguntas directas con respuestas “sí” o “no”.

La única pregunta fue: “¿Se lava las manos con regularidad y durante el tiempo suficiente (al menos 20 segundos) con agua y jabón?”

Este cuestionario obtuvo una muestra  $n=490$ .

- **La encuesta indirecta:** se divide en dos subgrupos:

**Grupo Modelo Crosswise 1**, con estas preguntas:

“1. ¿El cumpleaños de su madre es en mayo, junio o julio?”

2. ¿Se lava las manos con regularidad y durante el tiempo suficiente (al menos 20 segundos) con agua y jabón? “

Las respuestas posibles eran: "Mi respuesta es 'sí' a ambas preguntas o 'no' a ambas preguntas" o "Mi respuesta es 'sí' a una pregunta y 'no' a la otra pregunta (independientemente de cuál sea)".

Este subgrupo obtuvo una muestra de  $n=471$ .

**Grupo Modelo Crosswise 2**, con estas preguntas:

“1. ¿El cumpleaños de su madre es en agosto, septiembre, octubre, noviembre, diciembre, enero, febrero, marzo o abril?”

2. ¿Se lava las manos con regularidad y durante el tiempo suficiente (al menos 20 segundos) con agua y jabón? “

Las respuestas posibles eran: "Mi respuesta es 'sí' a ambas preguntas o 'no' a ambas preguntas" o "Mi respuesta es 'sí' a una pregunta y 'no' a la otra pregunta (independientemente de cuál sea)".

Este subgrupo obtuvo una muestra de  $n=472$ .

## Resultados.

En primer lugar se tuvo en cuenta las estadísticas del nacimiento de las madres entre mayo y junio, provistas por la Agencia Federal Alemana de Estadística (Pöttsch, 2012), la cual es  $p = 0,26$ . Conociendo este dato, y sabiendo que  $\pi_1 = \pi_2$ ,  $G^2(1) = 0.41$ ,  $p = .524$  lo que demuestra una buena adherencia a la metodología, se puede aplicar el Método Crosswise de la forma en el que se ha explicado arriba.

En el cuestionario directo la estimación de que la muestra se lavaba las manos correcta y suficientemente fue del 94,5%. Sin embargo, el resultado en el cuestionario indirecto bajó a 78,1%. Una diferencia muy notable, del 16,4%.

Los autores de este estudio concluyen que la técnica indirecta del método crosswise, efectivamente ofrece un mayor alto grado de confidencialidad que las técnicas directas. Los cuestionarios de preguntas directas, cuando se trata de

cuestiones sensibles normalmente se ven influidos por el sesgo de deseabilidad social.

### 3. Aplicación a una encuesta real.

Ahora que ya se ha hablado de la Covid-19, de las medidas de prevención asociadas a la enfermedad, y de la técnica de estimación indirecta, el Método Crosswise, se va a ver la aplicación del presente trabajo de investigación, con su pertinente análisis.

#### 3.1. Objetivos del estudio.

La pandemia de coronavirus, surgida a finales del año 2019, es un hecho que pocas veces a lo largo de la historia de la humanidad sucede. Enfermedades de esta magnitud deben ser objeto de estudio, pues todo lo que se pueda aprender servirá para que la sociedad en un futuro pueda estar mucho más preparada para este tipo de situaciones. Cuanto más se sepa, más preparación en el futuro, más igualdad, y mejor relación de las personas con el planeta habrá.

La complejidad de la Covid-19 hace pensar a los expertos que es muy probable que la enfermedad conviva con los humanos de aquí en adelante, no se sabe hasta cuanto tiempo más. Esto plantea la cuestión de saber no sólo si las medidas tomadas por gobiernos y autoridades sanitarias son adecuadas, sino si la sociedad en general está dispuesta a llevarlas a cabo, por las dificultades y sacrificios que conlleva.

El incumplimiento de las medidas impuestas por las autoridades, es algo que socialmente es indeseable e incluso condenable. Por eso, cuando a alguien se le pregunta acerca de su compromiso con las medidas, la respuesta puede no ser del todo honesta, al aplicarse el, ya mencionado, sesgo de deseabilidad social.

De cara al futuro, es también necesario encontrar una forma en la que conseguir información veraz, que se corresponda con los hechos reales, con las opiniones y con lo que de verdad piensa la gente. Conseguir que las personas que participan en



cualquier tipo de estudio, perciban que la confidencialidad nunca se verá comprometida y que pueden ser totalmente honestos.

Los objetivos de este trabajo por tanto son dos:

- El primero es a nivel teórico. Se pretende verificar que, efectivamente, el Método Crosswise consigue solventar el problema del sesgo de deseabilidad social cuando se pregunta por temas comprometidos en los que la persona que participa se expone a ser juzgada por su comportamiento o condición. En este aspecto, cabe esperar que en el análisis de la encuesta indirecta, con respecto a la directa, se obtenga el resultado que indique que se llevaron a cabo las medidas de prevención impuestas en menor grado.
- El segundo, a nivel social. Se pretende conocer cómo se está llevando a cabo la lucha contra esta pandemia. Qué grado de compromiso existe para llevar a cabo las medidas. Si los gobiernos deben optar por otras vías en lugar de la vía de la sanción por incumplimiento. También se quiere saber cómo conseguir de forma efectiva información honesta por parte de la gente, para que realmente los estudios tengan validez y utilidad científica y social.

### 3.2. Universo y muestra.

El total de cuestionarios realizados recibidos ha sido 667. Todos son válidos, porque para poder completarlo y enviarlo es necesario que todas las preguntas sean contestadas, y sólo admiten opciones cerradas, ninguna pregunta es abierta, por lo que no hay margen de error para obtener respuestas perdidas o inválidas.

El cuestionario ha sido difundido por las listas de distribución de correo electrónico de la Universidad de Granada. Así como difundido por contactos que también forman parte de la UGR. Se utilizó un muestreo de voluntarios, no había ninguna condición que cumplir para poder realizar la encuesta, excepto la de haber vivido las

restricciones por la Covid-19 en España, para poder dar una opinión en base a la experiencia de ello, y estar matriculado en la UGR. El universo total era 64.622 estudiantes (Secretaría General UGR, 2020).

### 3.3. Metodología.

La metodología llevada a cabo, tanto para el diseño, como para el análisis de los datos ha sido la metodología de técnica de estimación indirecta denominada “Método Crosswise”. Una extensión, y por tanto mejora, del Modelo Warner como técnica de respuesta aleatorizada. En el marco teórico de este trabajo de investigación se encuentran ambos métodos explicados.

### 3.4. Fases del estudio y programación.

Este es el esquema de trabajo que se ha seguido para la realización del presente trabajo de investigación:

<b>Método Crosswise. Aplicación a una encuesta sobre el cumplimiento de las medidas contra la COVID-19</b>		<b>2021</b>																			
		<b>MARZO</b>				<b>JUNIO</b>				<b>JULIO</b>				<b>AGOSTO</b>				<b>SEPTIEMBRE</b>			
<b>Objetivos</b>	<b>Duración</b>	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
<b>Elección de tema de investigación</b>	<b>2 semanas</b>																				
Elección de tema	2 sem																				
Puesta en común con las tutoras	1 sem																				
Inicio del proyecto de investigación	-																				
<b>Elaboración del marco teórico</b>	<b>5 semanas</b>																				
Datos sobre la COVID-19	1 sem																				
Estudios sobre la COVID-19	2 sem																				
Teoría técnicas estimaciones indirectas	2 sem																				
<b>Diseño e implementación del trabajo de campo</b>	<b>9 semanas</b>																				
Diseño de la encuesta	1 sem																				
Publicación y difusión de la encuesta	7 sem																				
Cambios para mejorar el muestreo	2 sem																				
<b>Análisis de datos obtenidos</b>	<b>3 semanas</b>																				
Depuración de los datos	1 sem																				
Análisis/comparativa de las dos encuestas	2 sem																				
Análisis /explicación de los resultados	1 sem																				
<b>Fase final del proyecto</b>	<b>2 semanas</b>																				
Conclusiones del proyecto	1 sem																				
Depósito de trabajo de investigación	1 sem																				
Elaboración de soporte presentación	1 sem																				
Presentación final ante tribunal																					
<b>TIEMPO TOTAL</b>	<b>15 semanas</b>																				
<b>COMENTARIOS</b>		<p>1 Desde marzo hasta junio no es posible el comienzo del proyecto por problemas de agenda</p> <p>2 Constante revisión</p>																			

### 3.5. Diseños de las diferentes encuestas.

Se mostrarán ahora cada una de las encuestas y como están diseñadas. Cada pregunta que hay en las encuestas tiene su justificación en la información que hay en el marco teórico que abarca las dos primeras secciones de este trabajo, es decir, todas están formuladas en base a hipótesis que se pueden deducir de la información recabada.

#### 3.5.1. Encuesta directa.

Se adjunta ahora la encuesta directa. Previamente hay un bloque de preguntas sociodemográficas para conocer mejor la muestra, cuyos resultados se verán más adelante. Es una encuesta compuesta por preguntas con variables dicotómicas. Las únicas respuestas posibles son “Sí” o “No”.

<b>Cumplimiento de las medidas preventivas para evitar la transmisión del Coronavirus en España</b>	
Bienvenid@, Este cuestionario pretende conocer en mayor profundidad en qué medida se cumplió con las medidas para combatir la pandemia de SARS-CoV-2 en España. Se le pide que responda a todas las cuestiones con total sinceridad, pues el anonimato en las respuestas está totalmente garantizado.	
<b>Variables sociodemográficas.</b> Responda, por favor, a unas breves preguntas de clasificación poblacional.	
<b>1.</b>	<b>Sexo</b>
1.	Mujer
2.	Hombre
3.	Otro

99.	Prefiero no decirlo
<b>2.</b>	<b>Edad</b>
	<i>Introducir número de años cumplidos</i>
<b>3.</b>	<b>Nacionalidad</b>
1.	Española
2.	Otro país europeo
3.	País americano
4.	Otro
<b>4.</b>	<b>Situación profesional</b>
1.	Empleado/a (incluye autónomos, situación ERTE o bajas)
2.	Sin empleo (en búsqueda activa de trabajo)
3.	Sin empleo por decisión propia
4.	Jubilado/a o pensionista
5.	Otro
99.	NS/NC
<b>5.</b>	<b>Por favor, marque la última cifra numérica que contenga su DNI, NIE o número de pasaporte</b>  <i>Esta pregunta tiene como fin redirigir al encuestado hacia el cuestionario directo o indirecto, dependiendo de cual sea el número en el que termina su número de identidad, los números 2,4,6,8 redirigen al directo y los números 0,1,3,5,7,9 al indirecto. Más números redirigen al cuestionario indirecto porque se necesitan muestras mayores, como ya se ha visto anteriormente.</i>
1.	0
2.	1
3.	2
4.	3
5.	4
6.	5
7.	6

8.	7
9.	8
10.	9
<p><b>Cuestiones sobre las medidas del gobierno contra la COVID-19</b>  Responda a las siguientes cuestiones, por favor. Recuerde que esta encuesta es totalmente anónima.  <i>(Comienzo del cuestionario directo)</i></p>	
1.	<b>¿Ha cumplido de forma estricta la medida de llevar mascarilla puesta fuera de su domicilio, en todo momento, desde el momento en que el gobierno la estableció obligatoria?</b>
1.	Sí.
2.	No.
2.	<b>¿Ha cumplido de forma estricta la medida de lavarse frecuentemente las manos o cada vez que entrara en un establecimiento público, con jabón y agua o gel hidroalcohólico, desde el momento en el que el gobierno la estableció obligatoria?</b>
1.	Sí.
2.	No.
3.	<b>¿Ha cumplido de forma estricta el confinamiento domiciliario que fue establecido por el Gobierno durante el Estado de Alarma, saliendo del domicilio sólo por razones de primera necesidad tales como hacer la compra o una emergencia médica? Periodo referido al confinamiento estricto que se llevó a cabo desde el 15 de marzo hasta el 21 de junio de 2020</b>
1.	Sí.
2.	No.
4.	<b>¿Ha cumplido de forma estricta la medida de mantener una distancia de seguridad por la calle y en establecimientos públicos de 1,5 metros con personas que no fueran de su burbuja de convivencia?</b>
1.	Sí.
2.	No.
5.	<b>¿Ha cumplido de forma estricta la medida de respetar el confinamiento perimetral, ya sea municipal, comarcal, provincial o regional? Medida impuesta después del confinamiento que limitaba la movilidad entre diferentes áreas del país</b>

1.	Sí.
2.	No.
<b>6.</b>	<b>¿Ha cumplido la medida del toque de queda casa según el horario establecido por las autoridades sanitarias correspondientes en su territorio?</b>
1.	Sí.
2.	No.

Aquí finaliza el cuestionario directo.

Para que los dos cuestionarios puedan ser comparables, ambos deben ser con preguntas dicotómicas, de sólo dos categorías. Esto también hace al encuestado que se sitúe más cercano a la verdad, teniendo que elegir entre “Sí” o “No”, pues, en temas comprometidos, cuando se ofrecen opciones intermedias la muestra tiende a centralizarse en sus respuestas.

### 3.5.2. Encuesta indirecta.

Se va a exponer ahora la encuesta indirecta. Diseñada en base al Método Crosswise.

<b>Cumplimiento de las medidas preventivas para evitar la transmisión del Coronavirus en España</b>
Bienvenid@, Este cuestionario pretende conocer en mayor profundidad en qué medida se cumplió con las medidas para combatir la pandemia de SARS-CoV-2 en España. Se le pide que responda a todas las cuestiones con total sinceridad, pues el anonimato en las respuestas está totalmente garantizado.
<b>Variabes sociodemográficas.</b> Responda, por favor, a unas breves preguntas de clasificación poblacional.

<b>1.</b>	<b>Sexo</b>
1.	Mujer
2.	Hombre
3.	Otro
99.	Prefiero no decirlo
<b>2.</b>	<b>Edad</b>
	<i>Introducir número de años cumplidos</i>
<b>3.</b>	<b>Nacionalidad</b>
1.	Española
2.	Otro país europeo
3.	País americano
4.	Otro
<b>4.</b>	<b>Situación profesional</b>
1.	Empleado/a (incluye autónomos, situación ERTE o bajas)
2.	Sin empleo (en búsqueda activa de trabajo)
3.	Sin empleo por decisión propia
4.	Jubilado/a o pensionista
5.	Otro
99.	NS/NC
<b>5.</b>	<b>Por favor, marque la última cifra numérica que contenga su DNI, NIE o número de pasaporte</b>
1.	0
2.	1
3.	2
4.	3
5.	4
6.	5
7.	6



8.	7
9.	8
10.	9
<p><b>Cuestiones sobre las medidas del gobierno contra la COVID-19</b></p> <p>Ahora se le presentarán diferentes bloques de preguntas los cuales tiene que responder en conjunto, es decir, señalando si responde SÍ o NO a ambas preguntas, o por el contrario, una de ellas es SÍ y otra es NO, independientemente cuál sea afirmativa y cuál negativa, no importa el orden. Este tipo encuesta tiene un procedimiento de pregunta indirecta, el cual garantiza que el entrevistador/a no pueda conocer su respuesta. (Comienzo del cuestionario indirecto)</p>	
<b>BLOQUE 1.</b>	<ul style="list-style-type: none"> <li>· <b>PREGUNTA 1A: ¿Ha cumplido de forma estricta la medida de llevar la mascarilla puesta fuera de su domicilio, en todo momento, desde el momento en que el Gobierno la estableció obligatoria?</b></li> <li>· <b>PREGUNTA 1B: ¿Ha nacido entre los meses de enero, febrero, marzo, o abril?</b></li> </ul>
1.	SÍ a las dos preguntas o NO a las dos preguntas.
2.	SÍ a una de las preguntas y NO a otra de las preguntas (sea el orden que sea).
<b>BLOQUE 2.</b>	<ul style="list-style-type: none"> <li>· <b>PREGUNTA 2A: ¿Ha cumplido de forma estricta la medida de lavarse frecuentemente las manos o cada vez que entrara en un establecimiento público, con jabón y agua o gel hidroalcohólico, desde el momento en que el Gobierno la estableció obligatoria?</b></li> <li>· <b>PREGUNTA 2B: ¿Ha finalizado los estudios secundarios obligatorios?</b></li> </ul>
1.	SÍ a las dos preguntas o NO a las dos preguntas.
2.	SÍ a una de las preguntas y NO a otra de las preguntas (sea el orden que sea).
<b>BLOQUE 3.</b>	<ul style="list-style-type: none"> <li>· <b>PREGUNTA 3A: ¿Ha cumplido de forma estricta el confinamiento domiciliario que fue establecido por el Gobierno durante el Estado de Alarma, saliendo del domicilio sólo por razones de primera necesidad, tales como hacer la compra o una emergencia médica? Periodo referido al confinamiento estricto que se llevó a cabo desde el 15 de marzo hasta el 21 de junio de 2020</b></li> <li>· <b>PREGUNTA 3B: ¿Su número de teléfono móvil termina en 0,</b></li> </ul>

	<b>1, 2, 3, 4, 5 o 6?</b>
1.	SÍ a las dos preguntas o NO a las dos preguntas.
2.	SÍ a una de las preguntas y NO a otra de las preguntas (sea el orden que sea).
<b>BLOQUE 4.</b>	<ul style="list-style-type: none"> <li>· <b>PREGUNTA 4A: ¿Ha cumplido de forma estricta la medida de mantener una distancia de seguridad por la calle y en establecimientos públicos de 1'5 metros con personas que no fueran de su burbuja de convivencia?</b></li> <li>· <b>PREGUNTA 4B: ¿Tiene usted licencia de conducción en España? Referido al carnet de coche estándar (B)</b></li> </ul>
1.	SÍ a las dos preguntas o NO a las dos preguntas.
2.	SÍ a una de las preguntas y NO a otra de las preguntas (sea el orden que sea).
<b>BLOQUE 5.</b>	<ul style="list-style-type: none"> <li>· <b>PREGUNTA 5A: ¿Ha cumplido de forma estricta la medida de respetar el confinamiento perimetral, ya sea municipal, comarcal, provincial o regional? Medida impuesta después del confinamiento domiciliario que limitaba la movilidad entre diferentes áreas del país.</b></li> <li>· <b>PREGUNTA 5B: ¿El cumpleaños de su madre es entre las fechas 1 de Enero y 30 de Abril (inclusive)?</b></li> </ul>
1.	SÍ a las dos preguntas o NO a las dos preguntas.
2.	SÍ a una de las preguntas y NO a otra de las preguntas (sea el orden que sea).
<b>BLOQUE 6.</b>	<ul style="list-style-type: none"> <li>· <b>PREGUNTA 6A: ¿Ha cumplido de forma estricta la medida del toque de queda en casa según el horario establecido por las autoridades sanitarias correspondientes en su territorio?</b></li> <li>· <b>PREGUNTA 6B: ¿Su cumpleaños es entre los días 1 al 11 del mes?</b></li> </ul>
1.	SÍ a las dos preguntas o NO a las dos preguntas.
2.	SÍ a una de las preguntas y NO a otra de las preguntas (sea el orden que sea).

Aquí se presentan las prevalencias necesarias que se han de saber de cada pregunta no sensible de cada bloque:

Bloque	Dato	Estimación	Fuente
1	Nacer entre los meses de enero/abril.	$p = 4/12 = 0,33$ $1 - p = 0,67$	(Cálculo manual)
2	Número de personas con los estudios secundarios finalizados	$p = 0,63$ $1 - p = 0,37$	Ministerio de Educación, Cultura y Desarrollo, 2020.
3	Número de teléfono móvil acabado en números del 0 al 6	$p = 7/10 = 0,70$ $1 - p = 0,30$	(Cálculo manual)
4	Número de personas con el Carnet B en España	$p = 0,57$ $1 - p = 0,43$	Ministerio del Interior, Dirección General de Tráfico, 2020.
5	Cumpleaños de la madre del 1 de enero al 30 de abril	$p = 120/365 = 0,32$ $1 - p = 0,68$	(Cálculo manual)
6	Cumpleaños encuestado entre los días 1 al 11	$p = 11/30 = 0,37$ $1 - p = 0,63$	(Cálculo manual)

Tabla estimaciones resultados preguntas no sensibles. Elaboración propia con datos de diversas fuentes, especificadas en la tabla.

## 4. Análisis.

### 4.1. Depuración de datos y pasos previos al análisis.

Una vez obtenida la matriz de respuestas, hay que proceder a la codificación, que se realizó de la siguiente manera:

Bloque de preguntas	Pregunta	Valores
	ID CUESTIONARIO	Valor de marca temporal para identificar el cuestionario, se sustituye por número consecutivos del 1 al 667, así se otorga a cada cuestionario un número único.
Sociodemográfico	SEXO	1=Mujer 2=Hombre 3=Otro 99=Prefiero no decirlo
Sociodemográfico	EDAD	Edad en valor numérico
Sociodemográfico	NACIONALIDAD	1=Española 2=Otro País europeo 3=País americano 4=Otro
Sociodemográfico	SITUACIÓN LABORAL	1=Empleado/a (incluye autónomos, situación ERTE o bajas) 2=Sin empleo (en búsqueda activa de trabajo) 3=Sin empleo por decisión propia 4=Jubilado/a o pensionista 5=Otro 99=NS/NC
	ID DNI	Última cifra del documento de identidad pertinente en valor numérico
	ID MÉTODO	Esta variable se crea para poder hacer una submuestra y diferenciar entre lo que han contestado el cuestionario directo y los que han contestado el indirecto.

Cuestionario Directo	P1 a P6	0=No 1=Sí
Cuestionario Indirecto	P1 a P6	0=Sí/No o No/Sí a las preguntas 1=Sí/Sí o No/No a ambas preguntas

Tabla: Codificación de las preguntas de los cuestionarios.

En cuanto a la depuración de datos no fue necesario hacer ningún cambio, pues las respuestas eran todas obligatorias y cerradas, lo que no daba lugar a respuestas sin contestar o contestadas con dato incorrecto.

## 4.2. Composición de la muestra y sus características.

Como ya se ha mencionado el total de los cuestionarios recibidos ha sido 667.

Se va a detallar ahora cómo se estructura la muestra en base a cada pregunta del bloque sociodemográfico:

- Sexo:

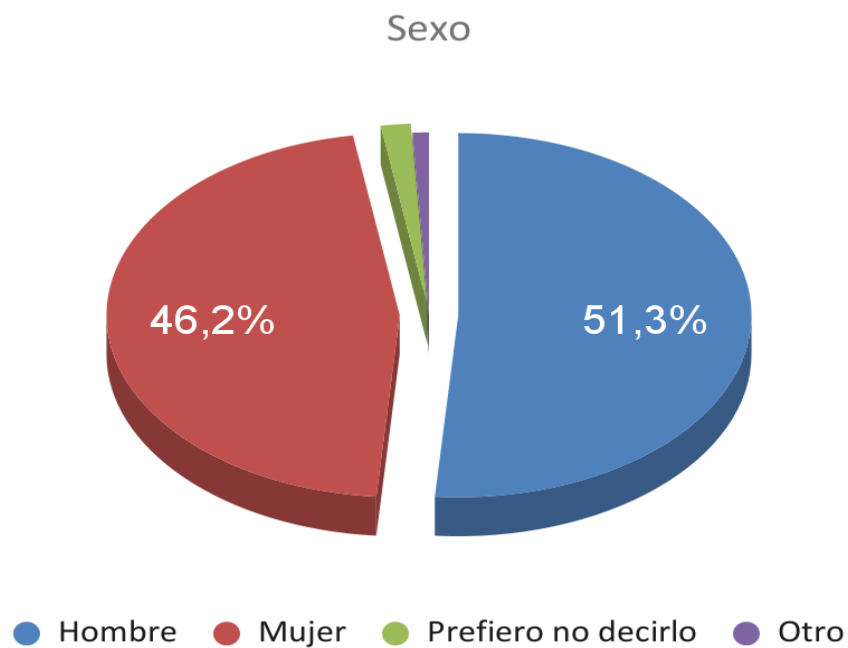


Gráfico: composición de la muestra por sexos.

Sexo	Porcentaje
Hombre	51,27%
Mujer	46,18%
Otro	0,90%
Prefiero no decirlo	1,65%
<b>Total</b>	<b>100,00%</b>

Tabla: composición de la muestra por sexos.

Como se puede observar, la muestra está bastante equilibrada en cuanto a sexo de los encuestados. 51,27% se consideran hombres y 46,18% se consideran mujeres. Tan sólo 2,55% de la muestra se identifica (o no ha querido hacerlo) con el género binario hombre-mujer.

- Edad:

<i>Edad</i>	<i>Porcentaje</i>
<b>17</b>	1,75%
<b>18</b>	6,14%
<b>19</b>	3,29%
<b>20</b>	3,51%
<b>21</b>	3,29%
<b>22</b>	5,48%
<b>23</b>	4,17%
<b>24</b>	8,55%
<b>25</b>	4,61%
<b>26</b>	3,73%
<b>27</b>	6,58%
<b>28</b>	9,43%
<b>29</b>	8,33%
<b>30</b>	9,87%
<b>31</b>	5,92%
<b>32</b>	5,48%
<b>33</b>	4,17%
<b>34</b>	3,51%
<b>35</b>	2,19%
<b>Total</b>	<b>100,00%</b>

Tabla: composición de la muestra por edad.

En este caso, las edades de los universitarios/as con mayor número de registro son: 30, 28, 24, 29, 27 y 18 años. Agrupando entre los 24 y 30 un grupo de edad con representación importante.

- Nacionalidad:

## Nacionalidad

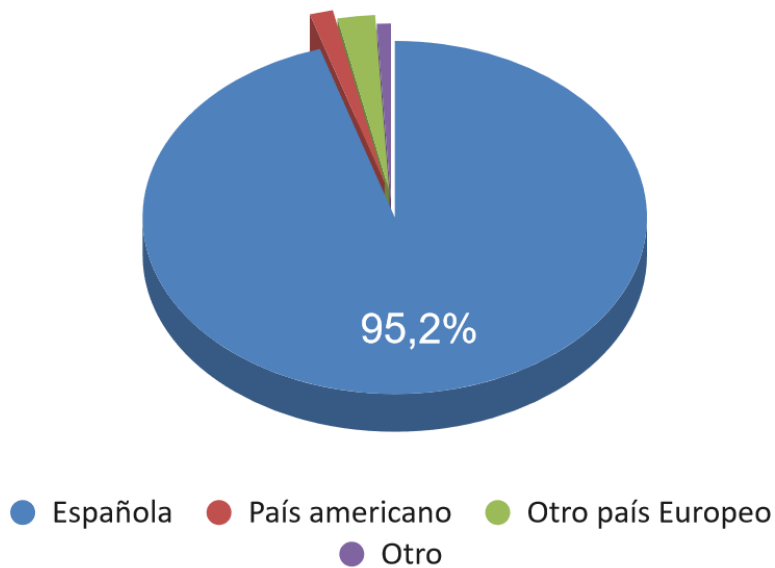


Gráfico: composición de la muestra según nacionalidad.

Nacionalidad	Porcentaje
Española	95,20%
Otro país Europeo	2,40%
País americano	1,50%
Otro	0,90%
<b>Total</b>	<b>100,00%</b>

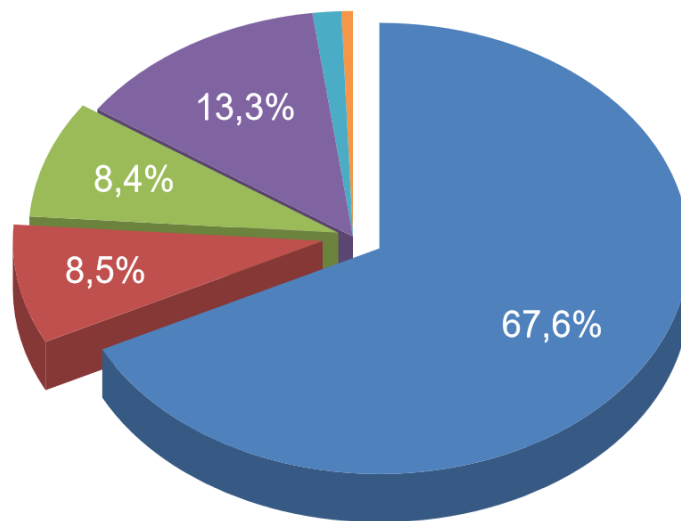
Tabla: composición de la muestra según nacionalidad.

La gran mayoría de la muestra es de nacionalidad española (95,20%) lo cual es bueno para la obtención de datos, pues se pregunta por el cumplimiento de las medidas adoptadas por el Gobierno de España, esto puede ofrecer un mejor conocimiento y una respuesta más fundada.

- Situación laboral:



### Situación laboral



- Empleado/a (incluye autónomos, situación ERTE o bajas)
- Sin empleo por decisión propia
- Sin empleo (en búsqueda activa de trabajo)
- Otro
- Jubilado/a o pensionista
- NS/NC

Gráfico: composición de la muestra por su situación laboral.

<i>Situación profesional</i>	Porcentaje
Empleado/a (incluye autónomos, situación ERTE o bajas)	67,62%
Otro	13,34%
Sin empleo por decisión propia	8,55%
Sin empleo (en búsqueda activa de trabajo)	8,40%
Jubilado/a o pensionista	1,50%
NS/NC	0,60%
<b>Total</b>	<b>100,00%</b>

Tabla: composición de la muestra por su situación laboral.

Aquí sí se puede encontrar una variedad en la distribución de la muestra. Un 67,62% se atribuye con trabajo, aunque esto también puede incluir estudiar a la vez, o al menos estar matriculado en la Universidad de Granada, también puede incluirse aquí el personal docente de la UGR. Las categorías, “Otro” y “Sin empleo por decisión propia”, las cuales juntas suman el 21,89% de la muestra, se atribuyen a

gente que únicamente estudia o por algún otro motivo decide no trabajar. Luego tenemos las opciones de “búsqueda activa de empleo” y de personas que cobran pensión con porcentajes menos significativos.

## **CREACIÓN DE SUBMUESTRAS**

Para un correcto análisis, fue necesario crear submuestras en base a dos grupos, teniendo en cuenta como característica el haber realizado el cuestionario directo o el indirecto. Por lo que se acaba dividiendo la muestra entre:

- Grupo Directo: n=278.
- Grupo Crosswise: n=389.

Se ha tenido que hacer esta separación porque la matriz de datos ha llegado toda junta. Recordemos que el filtro para hacer un cuestionario u otro es la última letra del número del documento de identidad. En base a esta separación se han realizado las dos submuestras.

La creación de las submuestras es elemental para el análisis en R de las estimaciones de cada grupo, y así poder compararlas.

### **4.3. Análisis de los datos.**

El análisis se ha llevado a cabo con el software estadístico R. Dentro de R se han utilizado los paquetes “RRreg” (Heck & Moshagen, 2018) y “sampling” (Tillé & Matei, 2021). El primero sirve para realizar obtener las estimaciones de las preguntas construidas en base al modelo Crosswise. El segundo para obtener las estimaciones y varianzas de las preguntas directas, lo que permitirá hacer contraste entre ambas.

## 4.4. Resultados.

### Estimaciones cuestionario directo:

Primero se presentarán las estimaciones de las preguntas realizadas de forma directa.

### Resultados generales

	Estimación	Varianza
Pregunta 1	0.6203209	0.001262588
Pregunta 2	0.6417112	0.001232541
Pregunta 3	0.8235294	0.0007790762
Pregunta 4	0.4545455	0.001329121
Pregunta 5	0.6470588	0.001224263
Pregunta 6	0.7433155	0.001022825

Tabla: resultados generales de las preguntas del cuestionario directo.

Según los resultados del cuestionario directo las medidas se han cumplido no muy a rajatabla, pues rondan los valores del 60% en las preguntas 1, 2 y 5, y baja hasta el 45% en la 4. Las medidas que parecen haberse cumplido más han sido el confinamiento (p3) y el toque de queda (p6).

### Resultados segmentados

- Sexo:

Sexo	Estimación		p valor
	Mujeres	Hombres	
Pregunta 1	0.6452	0.587	0.124

<b>Pregunta 2</b>	0.7419	0.5326	0.0478
<b>Pregunta 3</b>	0.8925	0.7609	0.0253
<b>Pregunta 4</b>	0.4839	0.4239	0.4661
<b>Pregunta 5</b>	0.6989	0.587	0.0917
<b>Pregunta 6</b>	0.7634	0.7174	0.6715

Tabla: resultados por sexo de las preguntas del cuestionario directo.

En general, las mujeres obtienen puntuaciones más altas que los hombres en el cumplimiento de todas las medidas.

- Nacionalidad:

Hay que aclarar que la muestra se componía en más de un 95% de población española. Los valores de las otras nacionalidades son poco representativos, por eso no se exponen aquí. Por otro lado, los valores de la población española son bastante similares a los resultados generales.

- Situación laboral:

Algo similar ocurre con la muestra de jubilados, es muy pequeña y no es representativa.

Estimaciones cuestionario indirecto:

### **Resultados generales**

	<b>Estimación</b>	<b>Varianza</b>
<b>Pregunta 1</b>	0.798625	0.089302
<b>Pregunta 2</b>	1.000	2e-06
<b>Pregunta 3</b>	0.859195	0.074252
<b>Pregunta 4</b>	0.78736	0.22077
<b>Pregunta 5</b>	0.654321	0.085602
<b>Pregunta 6</b>	0.50737	0.11926

Tabla: resultados generales de las preguntas del cuestionario indirecto.

Los resultados obtenidos a través del cálculo con el paquete RRreg para el método Crosswise nos devuelven unas estimaciones mayores que los resultados del método directo. Lo cual es lo opuesto a la hipótesis que se esperaba cumplir. Más adelante se comentará. Se exponen todos los datos segmentados por las variables demográficas para que se puedan ver en detalle.

### Resultados segmentados

- Sexo:

Sexo	Estimación		p valor
	Mujeres	Hombres	
Pregunta 1	0.88309	0.74144	0.1590
Pregunta 2	1.0000	1.0000	0.9741
Pregunta 3	0.99370	0.72388	0.0667
Pregunta 4	0.77011	0.65992	0.2610
Pregunta 5	0.0091	0.68657	0.0513
Pregunta 6	0.71008	0.38519	0.0571

Tabla: resultados por sexo de las preguntas del cuestionario indirecto.

Si consideramos como nivel de significación 0.1, encontramos diferencias significativas en las preguntas 3, 5 y 6. Donde podemos considerar la estimación entre hombres y mujeres distinta.

- Nacionalidad: como antes, tampoco se incluyen por no ser significativos y ser la gran mayoría de los participantes españoles.
- Situación laboral: poca representación también.

Comparativas y contrastes:

	<b>Estimación Directa</b>	<b>Estimación Indirecta</b>	<b>p-valor</b>
<b>Pregunta 1</b>	0.6203209	0.798625	0.064
<b>Pregunta 2</b>	0.6417112	1.000	0.000
<b>Pregunta 3</b>	0.8235294	0.859195	0.653
<b>Pregunta 4</b>	0.4545455	0.78736	0.137
<b>Pregunta 5</b>	0.6470588	0.654321	0.937
<b>Pregunta 6</b>	0.7433155	0.50737	0.056

Tabla: comparación de resultados directos e indirectos y coeficiente de correlación.

Ahora que tenemos una imagen comparativa en esta última tabla, se puede ver mejor que las estimaciones indirectas son, en general, mayores que las directas. Esperábamos justo lo contrario. Estos resultados nos indican que la gente ha cumplido en mayor grado las medidas cuando se pregunta mediante el método crosswise que cuando se pregunta de forma directa. El p-valor no indica evidencia estadística de que haya diferencia significativa entre las estimaciones directas e indirectas, a excepción de las preguntas 1 y 6. No queda demostrado que se haya solventado el sesgo de deseabilidad social, o que siquiera exista en el tema que estamos preguntando.

Para el contraste se ha utilizado la propuesta de Wolter & Preisendörfer (2013) los cuales proponen el siguiente contraste:

$$Z = \frac{Est\_directa - est\_indirecta}{\sqrt{var\_directa + var\_indirecta}}$$

El código también está recogido en el anexo.

Se ha realizado una regresión logística con los resultados del cuestionario indirecto y las variables sociodemográficas. Los resultados obtenidos indican que ninguna de las variables introducidas son estadísticamente significativas, por lo que no lo incluimos en la memoria, pero se adjunta el código utilizado al final del anexo, utilizando la función `RRlog` del paquete `RRreg`.

## 5. Conclusiones.

Teniendo los precedentes de Mieth & Et al. (2020) se esperaban unos resultados en los que el método crosswise nos desvelara que, preguntando de forma indirecta, se obtendría unas estimaciones más bajas de grado de cumplimiento con las medidas para frenar el coronavirus. Se encuentran varias explicaciones para este fenómeno:

### Aspectos técnicos:

- Es posible que la muestra haya sido insuficiente (667) para hacer una buena estimación y contrastes, se hubiera deseado que la muestra hubiese superado los n=1000 participantes, sobre todo en las personas que han realizado el cuestionario indirecto, que se hubiera deseado al menos el doble de muestra.
- Adherencia y comprensión del cuestionario por parte de la muestra. Es posible que los participantes no hayan comprendido del todo las instrucciones del cuestionario y cómo contestarlo. De forma que las respuestas pueden no corresponderse con la realidad.
- Sesgo de deseabilidad inexistente. Es posible que el tema no sea tan controvertido y la gente no necesite mentir u ocultar acerca de si han llevado a cabo o no las medidas de prevención contra el coronavirus.

### Aspectos sociales y contextuales:

- Momento de realización del estudio. Comparado con el estudio de Mieth & Et al. (2020), la pandemia está ahora mismo en un momento muy diferente. En ese momento (abril de 2020) no se había desarrollado vacuna alguna, estábamos en pleno confinamiento domiciliario y la pandemia era foco de todas las noticias que veíamos en los medios. Por lo tanto, la obligación de



las medidas de contención contra la COVID-19 eran totalmente obligatorias y llevaban consigo una gran carga de sesgo de deseabilidad social. Ahora estamos en un momento de mucha más apertura, relajación y tranquilidad, en el que no nos importa reconocer si no lo hicimos tan bien en aquella época tan dura. Y la vacunación de gran parte de la población contribuye mucho a esta tranquilidad.

- Información y desinformación. Si algo ha tenido esta pandemia ha sido el conflicto entre las diferentes informaciones que nos han ido llegando. Desde bulos hasta confusión entre las cifras oficiales. Esto ha despertado actitudes de incredulidad ante la verdadera existencia de la enfermedad, y gente que no ha llevado a cabo las medidas de contención, pues pueden no ser tan percibidas como algo que si no se cumple irá en perjuicio de toda la sociedad. Incluso hemos visto como gente ha adoptado posturas tan radicales como los negacionistas.
- Quizá el no cumplimiento de las medidas no sea un tema tan cargado de sesgo de deseabilidad social como otros como pueden ser la violencia de género, aborto, consumo de prostitución, etc.

## 6. Bibliografía.

- Andreu Cabrera, E. (2020). Actividad física y efectos psicológicos del confinamiento por covid-19. *Revista INFAD de Psicología*, 2(1), 209-220.  
<https://revista.infad.eu/index.php/IJODAEP/article/view/1828>
- Canal Sur. (2021, 05 09). *Cronología del estado de alarma: del confinamiento duro a la apertura casi total*. Canal Sur. Radio y Televisión.  
<https://www.canalsur.es/noticias/cronologia-del-estado-de-alarma-del-confinamiento-duro-a-la-apertura-casi-total/1712353.html>
- Centro de Investigaciones Sociológicas. (2020, 11 23). 3302| *EFFECTOS Y CONSECUENCIAS DEL CORONAVIRUS (II)*. CIS · Ficha del estudio.  
[http://www.cis.es/cis/opencm/ES/1\\_encuestas/estudios/ver.jsp?estudio=14534](http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=14534)
- Centro de Investigaciones Sociológicas. (2020, 12 11). 3305| *EFFECTOS Y CONSECUENCIAS DEL CORONAVIRUS (III)*. CIS · Ficha del estudio.  
[http://www.cis.es/cis/opencm/ES/1\\_encuestas/estudios/ver.jsp?estudio=14538](http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=14538)
- Centro de Investigaciones Sociológicas. (2021, 05 14). 3324| *EFFECTOS Y CONSECUENCIAS DEL CORONAVIRUS (IV)*. CIS · Ficha del estudio.  
[http://www.cis.es/cis/opencm/ES/1\\_encuestas/estudios/ver.jsp?estudio=14570](http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=14570)
- Cobo Rodríguez, B. (2013, 06). *Respuesta aleatoria y técnicas de pregunta indirectas* [Trabajo de investigación fin de máster]. Universidad de Granada.
- Devore, J. L. (1977). A note on the randomized response technique. *Communications in Statistics Theory and Methods*, 6, 1525-1529.

- Dirección General de Tráfico. (2021). *Censo de Conductores*. Tablas estadísticas. <https://www.dgt.es/es/seguridad-vial/estadisticas-e-indicadores/censo-conductores/tablas-estadisticas/>
- Gabinete de la presidencia del Gobierno. (2020, 03 04). *Coronavirus (COVID-19) - 04 de marzo de 2020. Situación actual*. Coronavirus (COVID-19) 04 marzo 2020 | DSN. <https://www.dsn.gob.es/es/actualidad/sala-prensa/coronavirus-covid-19-04-marzo-2020>
- Gimenez Luque, M. (2020). *Métodos de entrevista de preguntas sensibles: el método Crosswise. Aplicación a una encuesta sobre violencia de género* [Trabajo Fin de Grado]. Universidad de Granada.
- Greenberg, B. G., Abul-Ela, A. L., Simons, W. R., & Horvitz, D. G. (1969). The unrelated question RR model: Theoretical framework. *JASA*, *64*, 520-539.
- Heck, D. W., Hoffman, A., & Moshagen, M. (2018). Detecting nonadherence without loss in efficiency: A simple extension of the crosswise model. *Behavior Research Methods*, *50*(1), 1895-1905. <https://doi.org/10.3758/s13428-017-0957-8>
- Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The unrelated question RR model. *Proc. Social Statist. Sec. ASA*, 65-72.
- Instituto Nacional de Estadística. (2021, 01 1). *Población residente en España*. INEbase. [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176951&menu=ultiDatos&idp=1254735572981](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176951&menu=ultiDatos&idp=1254735572981)
- Johns Hopkins. University of Medicine. (2021, 07 30). *Coronavirus Tracker Map*. Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html>

La Moncloa. Gobierno de España. (2020, 03 13). *El Gobierno declarará mañana el estado de alarma por el coronavirus*. La Moncloa.

<https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2020/130320-sanchez-declaracio.aspx>

Mieth, L., Mayer, M., Hoffmann, A., Buchner, A., & Bell, R. (2020, 12 22). Do they really wash their hands? Prevalence estimates for personal hygiene behaviour during the COVID-19 pandemic based on indirect questions. *Research Square*. 10.21203/rs.3.rs-72009/v2

Ministerio de Educación y Formación Profesional. (2020, 06). *EXPLORACIÓN DE LAS VARIABLES EDUCATIVAS DE LA ENCUESTA DE POBLACIÓN ACTIVA / NIVEL DE FORMACIÓN DE LA POBLACIÓN*. EDUCAbase. <http://estadisticas.mecd.gob.es/EducaJaxiPx/Tabla.htm?path=/laborales/epa/nivfor/I0/&file=nivfor101.px&type=pcaxis&L=0>

Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. (2020, 03 14). *Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19*. Agencia Estatal Boletín Oficial del Estado. <https://www.boe.es/buscar/doc.php?id=BOE-A-2020-3692>

Ministerio de Sanidad, Consumo y Bienestar Social. (2021, 02 17). *Información para la ciudadanía. Las 6M siempre en mente*. Centro de Coordinación de Alertas y Emergencias Sanitarias (CCAES). <https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/ciudadania.htm>

- Organización Mundial de la Salud. (2020, 04 27). *COVID-19: cronología de la actuación de la OMS*. Organización Mundial de la Salud. Comunicados de prensa. <https://www.who.int/es/news/item/27-04-2020-who-timeline---covid-19>
- Perri, P. F., Pelle, E., & Stranges, M. (2015). Estimating Induced Abortion and Foreign Irregular Presence Using the Randomized Response Crossed Model. *Soc Indic Res*, 129. 10.1007/s11205-015-1136-x
- Person, B., Sy, F., Holton, K., Govert, B., Liang, A., & The NCID, SARS Community Outreach Team. (2004, 02). Fear and Stigma: The Epidemic within the SARS Outbreak. *Emerging Infectious Diseases*, 2(10), 358-363. National Center for Biotechnology Information. 10.3201/eid1002.030750
- Rodríguez-Parrales, D. H., Alegría-Caminos, Y. F., & Angulo-Corozo, J. I. (2020, 06 23). Epidemias causadas por coronavirus: Sintomatología clínica y hallazgos de laboratorio. *Dominio de las ciencias. Revista científica.*, 7(3), 710-744. <http://dx.doi.org/10.23857/dc.v7i3.2021>
- R. Quah, S., & Hin-Peng, L. (2004, 02). Crisis Prevention and Management during SARS Outbreak, Singapore. *Emerging infectious Diseases*, 10(2), 364–368. National Center for Biotechnology Information. 10.3201/eid1002.030418
- Ruiz-Bravo López, A., & Jiménez Valera, M. (2020, 03 30). SARS-CoV-2 y pandemia de síndrome respiratorio agudo (COVID-19). *Ars Pharm*, 61(2), 63-79. <http://dx.doi.org/10.30827/ars.v61i2.15177>
- Secretaría General UGR. (2020, 09 15). *Memoria Académica 2019-2020*. Secretaría General > Memoria Académica 2020. <https://secretariageneral.ugr.es/pages/memorias/academica/20192020>
- Statista. (2021, 09 03). *Número total de casos confirmados de coronavirus en España entre el 25 de febrero de 2020 y el 3 de septiembre de 2021*. Statista.

<https://es.statista.com/estadisticas/1104275/casos-confirmados-de-covid-19-por-dia-espana/>

- Tian, G.-L., & Tang, M.-L. (2014). *Incomplete Categorical Data Design*. Taylor & Francis Group.
- Tillé, Y., & Matei, A. (2021, 13 1). *sampling: survey sampling*. CRAN - Package sampling. <https://cran.r-project.org/package=sampling>
- Van Der Hout, A., Van Der Heijden, P., & Gilchrist, R. (1991). A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda. *Sociological Methods Research*, 20(1), 104-138.
- Warner, S. L. (1965). A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(1), 63-69.
- Wolter, F., & Preisendörfer, P. (2013). Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data. *Sociological Methods & Research*, 42(3), 321-353.

# ANEXOS.

## Código empleado en R-Studio.

```
#JAIME CASADO CHICA-ANÁLISIS ESTIMACIONES INDIRECTAS

library(readxl)
library(RRreg)
library(sampling)

-----

#CARGAR DATOS

datos <- read_excel("datos.xlsx")
View(datos)

-----

#forzar formato

datos$SEXO=as.factor(datos$SEXO)

datos$NACIONALIDAD=as.factor(datos$NACIONALIDAD)

datos$SITUAC_LAB=as.factor(datos$SITUAC_LAB)

datos$ID_METODO=as.factor(datos$ID_METODO)

#separar grupos directo e indirecto

grupodirecto=subset(poblacion_jovenes,poblacion_jovenes$ID_METODO==0)
grupocrosswise=subset(poblacion_jovenes,poblacion_jovenes$ID_METODO==1)

-----

#medias y varianzas

n0=278
n1=389
N=64622
pi=rep(n0/N,n0)

#1
HTP1DIRECTA<-HTestimator(grupodirecto$P1_DIRECTA_MASCARILLA,pik = pi)/N
VARP1DIRECTA_1<-varest(grupodirecto$P1_DIRECTA_MASCARILLA,pik = pi)/N^2
```

```

#2
HTP2DIRECTA<-HTestimator(grupodirecto$P2_DIRECTA_MANOS,pik = pi)/N
VARP2DIRECTA_2<-varest(grupodirecto$P2_DIRECTA_MANOS,pik = pi)/N^2

#3
HTP3DIRECTA<-HTestimator(grupodirecto$P3_DIRECTA_CONFINAMIENTO,pik = pi)/N
VARP3DIRECTA_3<-varest(grupodirecto$P3_DIRECTA_CONFINAMIENTO,pik = pi)/N^2

#4
HTP4DIRECTA<-HTestimator(grupodirecto$P4_DIRECTA_DISTANCIASEG,pik = pi)/N
VARP4DIRECTA_4<-varest(grupodirecto$P4_DIRECTA_DISTANCIASEG,pik = pi)/N^2

#5
HTP5DIRECTA<-HTestimator(grupodirecto$P5_DIRECTA_PERIMETRAL,pik = pi)/N
VARP5DIRECTA_5<-varest(grupodirecto$P5_DIRECTA_PERIMETRAL,pik = pi)/N^2

#6
HTP6DIRECTA<-HTestimator(grupodirecto$P6_DIRECTA_TOQUEQUEDA,pik = pi)/N
VARP6DIRECTA_6<-varest(grupodirecto$P6_DIRECTA_TOQUEQUEDA,pik = pi)/N^2

HTP1DIRECTA
HTP2DIRECTA
HTP3DIRECTA
HTP4DIRECTA
HTP5DIRECTA
HTP6DIRECTA
VARP1DIRECTA_1
VARP2DIRECTA_2
VARP3DIRECTA_3
VARP4DIRECTA_4
VARP5DIRECTA_5
VARP6DIRECTA_6

```

-----

#Estimaciones con método crosswise con el paquete "RRreg" y "función RRuni"

```

#1
crossp1=RRuni(response=grupocrosswise$P1_CROSSWISE_MASCARILLA,
              data=grupocrosswise, model="Crosswise",p=0.33)
summary(crossp1)

```

```

#2
crossp2=RRuni(response=grupocrosswise$P2_CROSSWISE_MANOS,

```



```

        data=grupocrosswise, model="Crosswise",p=0.63)
summary(crossp2)

#3
crossp3=RRuni(response=grupocrosswise$P3_CROSSWISE_CONFINAMIENTO,
              data=grupocrosswise, model="Crosswise",p=0.70)
summary(crossp3)

#4

crossp4=RRuni(response=grupocrosswise$P4_CROSSWISE_DISTANCIASEG,
              data=grupocrosswise, model="Crosswise",p=0.57)
summary(crossp4)

#5

crossp5=RRuni(response=grupocrosswise$P5_CROSSWISE_PERIMETRAL,
              data=grupocrosswise, model="Crosswise",p=0.32)
summary(crossp5)

#6

crossp6=RRuni(response=grupocrosswise$P6_CROSSWISE_TOQUEQUEDA,
              data=grupocrosswise, model="Crosswise",p=0.37)
summary(crossp6)

-----

#contrastes

#1

Z_p1<- (HTP1DIRECTA-crossp1$pi) /sqrt (VARP1DIRECTA_1+(crossp1$piSE)^2)

#2

Z_p2<- (HTP2DIRECTA-crossp2$pi) /sqrt (VARP2DIRECTA_2+(crossp2$piSE)^2)

#3

Z_p3<- (HTP3DIRECTA-crossp3$pi) /sqrt (VARP3DIRECTA_3+(crossp3$piSE)^2)

#4

Z_p4<- (HTP4DIRECTA-crossp4$pi) /sqrt (VARP4DIRECTA_4+(crossp4$piSE)^2)

#5

Z_p5<- (HTP5DIRECTA-crossp5$pi) /sqrt (VARP5DIRECTA_5+(crossp5$piSE)^2)

#6

```

```

Z_p6<- (HTP6DIRECTA-crossp6$pi) /sqrt (VARP6DIRECTA_6+ (crossp6$piSE)^2)

#P valor

#1
pvalorp1<-2*pnorm(abs(Z_p1),lower.tail=FALSE)

#2
pvalorp2<-2*pnorm(abs(Z_p2),lower.tail=FALSE)

#3
pvalorp3<-2*pnorm(abs(Z_p3),lower.tail=FALSE)

#4
pvalorp4<-2*pnorm(abs(Z_p4),lower.tail=FALSE)

#5
pvalorp5<-2*pnorm(abs(Z_p5),lower.tail=FALSE)

#6
pvalorp6<-2*pnorm(abs(Z_p6),lower.tail=FALSE)

pvalorp1
pvalorp2
pvalorp3
pvalorp4
pvalorp5
pvalorp6

-----

#segmentar

#sexo + univariante sexo

Mujeres_indirecto=subset(grupocrosswise,grupocrosswise$SEXO==1)
Hombres_indirecto=subset(grupocrosswise,grupocrosswise$SEXO==2)
Mujeres_directo=subset(grupodirecto,grupodirecto$SEXO==1)
Hombres_directo=subset(grupodirecto,grupodirecto$SEXO==2)

#1IND

crossp1_mujeres=RRuni(response=Mujeres_indirecto$P1_CROSSWISE_MASCARILLA,
                      data=Mujeres_indirecto, model="Crosswise",p=0.33)
summary(crossp1_mujeres)

crossp1_hombres=RRuni(Hombres_indirecto$P1_CROSSWISE_MASCARILLA,
                      data=Hombres_indirecto, model="Crosswise",p=0.33)
summary(crossp1_hombres)

```

```

#1DIR

summary(Mujeres_directo$P1_DIRECTA_MASCARILLA)
summary(Hombres_directo$P1_DIRECTA_MASCARILLA)

#2IND
crossp2_mujeres=RRuni(response=Mujeres_indirecto$P2_CROSSWISE_MANOS,
                      data=Mujeres_indirecto, model="Crosswise",p=0.63)
summary(crossp2_mujeres)

crossp2_hombres=RRuni(Hombres_indirecto$P2_CROSSWISE_MANOS,
                      data=Hombres_indirecto, model="Crosswise",p=0.63)
summary(crossp2_hombres)

#2DIR
summary(Mujeres_directo$P2_DIRECTA_MANOS)
summary(Hombres_directo$P2_DIRECTA_MANOS)

#3IND
crossp3_mujeres=RRuni(response=Mujeres_indirecto$P3_CROSSWISE_CONFINAMIENTO,
                      data=Mujeres_indirecto, model="Crosswise",p=0.70)
summary(crossp3_mujeres)

crossp3_hombres=RRuni(Hombres_indirecto$P3_CROSSWISE_CONFINAMIENTO,
                      data=Hombres_indirecto, model="Crosswise",p=0.70)
summary(crossp3_hombres)

#3DIR
summary(Mujeres_directo$P3_DIRECTA_CONFINAMIENTO)
summary(Hombres_directo$P3_DIRECTA_CONFINAMIENTO)

#4IND
crossp4_mujeres=RRuni(response=Mujeres_indirecto$P4_CROSSWISE_DISTANCIASEG,
                      data=Mujeres_indirecto, model="Crosswise",p=0.57)
summary(crossp4_mujeres)

crossp4_hombres=RRuni(Hombres_indirecto$P4_CROSSWISE_DISTANCIASEG,
                      data=Hombres_indirecto, model="Crosswise",p=0.57)
summary(crossp4_hombres)

#4DIR
summary(Mujeres_directo$P4_DIRECTA_DISTANCIASEG)
summary(Hombres_directo$P4_DIRECTA_DISTANCIASEG)

#5IND
crossp5_mujeres=RRuni(response=Mujeres_indirecto$P5_CROSSWISE_PERIMETRAL,
                      data=Mujeres_indirecto, model="Crosswise",p=0.532)
summary(crossp5_mujeres)

crossp5_hombres=RRuni(Hombres_indirecto$P5_CROSSWISE_PERIMETRAL,
                      data=Hombres_indirecto, model="Crosswise",p=0.32)
summary(crossp5_hombres)

```

```

#5DIR
summary(Mujeres_directo$P5_DIRECTA_PERIMETRAL)
summary(Hombres_directo$P5_DIRECTA_PERIMETRAL)

#6IND
crossp6_mujeres=RRuni(response=Mujeres_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                      data=Mujeres_indirecto, model="Crosswise",p=0.37)
summary(crossp6_mujeres)

crossp6_hombres=RRuni(Hombres_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                      data=Hombres_indirecto, model="Crosswise",p=0.37)
summary(crossp6_hombres)

#6DIR
summary(Mujeres_directo$P6_DIRECTA_TOQUEQUEDA)
summary(Hombres_directo$P6_DIRECTA_TOQUEQUEDA)

-----
--

#NACIONALIDAD + UNIVARIANTE NACIONALIDAD

espanya_indirecto=subset(grupocrosswise, grupocrosswise$NACIONALIDAD==1)
europa_indirecto=subset(grupocrosswise, grupocrosswise$NACIONALIDAD==2)
america_indirecto=subset(grupocrosswise, grupocrosswise$NACIONALIDAD==3)
otropais_indirecto=subset(grupocrosswise, grupocrosswise$NACIONALIDAD==4)
espanya_directo=subset(grupodirecto, grupodirecto$NACIONALIDAD==1)
europa_directo=subset(grupodirecto, grupodirecto$NACIONALIDAD==2)
america_directo=subset(grupodirecto, grupodirecto$NACIONALIDAD==3)
otropais_directo=subset(grupodirecto, grupodirecto$NACIONALIDAD==4)

#1IND
crossp1_espanya=RRuni(response=espanya_indirecto$P1_CROSSWISE_MASCARILLA,
                      data=espanya_indirecto, model="Crosswise",p=0.33)
summary(crossp1_espanya)

crossp1_europa=RRuni(europa_indirecto$P1_CROSSWISE_MASCARILLA,
                    data=europa_indirecto, model="Crosswise",p=0.33)
summary(crossp1_europa)

crossp1_america=RRuni(response=america_indirecto$P1_CROSSWISE_MASCARILLA,
                      data=america_indirecto, model="Crosswise",p=0.33)
summary(crossp1_america)

crossp1_otropais=RRuni(response=otropais_indirecto$P1_CROSSWISE_MASCARILLA,
                       data=otropais_indirecto, model="Crosswise",p=0.33)
summary(crossp1_otropais)

#1DIR
summary(espanya_directo$P1_DIRECTA_MASCARILLA)
summary(europa_directo$P1_DIRECTA_MASCARILLA)

```

```

summary(america_directo$P1_DIRECTA_MASCARILLA)
summary(otropais_directo$P1_DIRECTA_MASCARILLA)

#2IND
crossp2_espanya=RRuni(response=espanya_indirecto$P2_CROSSWISE_MANOS,
                      data=espanya_indirecto, model="Crosswise",p=0.63)
summary(crossp2_espanya)

crossp2_europa=RRuni(europa_indirecto$P2_CROSSWISE_MANOS,
                    data=europa_indirecto, model="Crosswise",p=0.63)
summary(crossp2_europa)

crossp2_america=RRuni(response=america_indirecto$P2_CROSSWISE_MANOS,
                      data=america_indirecto, model="Crosswise",p=0.63)
summary(crossp2_america)

crossp2_otropais=RRuni(response=otropais_indirecto$P2_CROSSWISE_MANOS,
                      data=otropais_indirecto, model="Crosswise",p=0.63)
summary(crossp2_otropais)

#2DIR
summary(espanya_directo$P2_DIRECTA_MANOS)
summary(europa_directo$P2_DIRECTA_MANOS)
summary(america_directo$P2_DIRECTA_MANOS)
summary(otropais_directo$P2_DIRECTA_MANOS)

#3IND
crossp3_espanya=RRuni(response=espanya_indirecto$P3_CROSSWISE_CONFINAMIENTO,
                      data=espanya_indirecto, model="Crosswise",p=0.70)
summary(crossp3_espanya)

crossp3_europa=RRuni(europa_indirecto$P3_CROSSWISE_CONFINAMIENTO,
                    data=europa_indirecto, model="Crosswise",p=0.70)
summary(crossp3_europa)

crossp3_america=RRuni(response=america_indirecto$P3_CROSSWISE_CONFINAMIENTO,
                      data=america_indirecto, model="Crosswise",p=0.70)
summary(crossp3_america)

crossp3_otropais=RRuni(response=otropais_indirecto$P3_CROSSWISE_CONFINAMIENTO
,
                      data=otropais_indirecto, model="Crosswise",p=0.70)
summary(crossp3_otropais)

#3DIR
summary(espanya_directo$P3_DIRECTA_CONFINAMIENTO)
summary(europa_directo$P3_DIRECTA_CONFINAMIENTO)
summary(america_directo$P3_DIRECTA_CONFINAMIENTO)
summary(otropais_directo$P3_DIRECTA_CONFINAMIENTO)

#4IND
crossp4_espanya=RRuni(response=espanya_indirecto$P4_CROSSWISE_DISTANCIASEG,
                      data=espanya_indirecto, model="Crosswise",p=0.57)

```

```

summary(crossp4_espanya)

crossp4_europa=RRuni(europa_indirecto$P4_CROSSWISE_DISTANCIASEG,
                    data=europa_indirecto, model="Crosswise",p=0.57)
summary(crossp4_europa)

crossp4_america=RRuni(response=america_indirecto$P4_CROSSWISE_DISTANCIASEG,
                    data=america_indirecto, model="Crosswise",p=0.57)
summary(crossp4_america)

crossp4_otropais=RRuni(response=otropais_indirecto$P4_CROSSWISE_DISTANCIASEG,
                    data=otropais_indirecto, model="Crosswise",p=0.57)
summary(crossp4_otropais)

#4DIR
summary(espanya_directo$P4_DIRECTA_DISTANCIASEG)
summary(europa_directo$P4_DIRECTA_DISTANCIASEG)
summary(america_directo$P4_DIRECTA_DISTANCIASEG)
summary(otropais_directo$P4_DIRECTA_DISTANCIASEG)

#5IND
crossp5_espanya=RRuni(response=espanya_indirecto$P5_CROSSWISE_PERIMETRAL,
                    data=espanya_indirecto, model="Crosswise",p=0.32)
summary(crossp5_espanya)

crossp5_europa=RRuni(europa_indirecto$P5_CROSSWISE_PERIMETRAL,
                    data=europa_indirecto, model="Crosswise",p=0.32)
summary(crossp5_europa)

crossp5_america=RRuni(response=america_indirecto$P5_CROSSWISE_PERIMETRAL,
                    data=america_indirecto, model="Crosswise",p=0.32)
summary(crossp5_america)

crossp5_otropais=RRuni(response=otropais_indirecto$P5_CROSSWISE_PERIMETRAL,
                    data=otropais_indirecto, model="Crosswise",p=0.32)
summary(crossp5_otropais)

#5DIR
summary(espanya_directo$P5_DIRECTA_PERIMETRAL)
summary(europa_directo$P5_DIRECTA_PERIMETRAL)
summary(america_directo$P5_DIRECTA_PERIMETRAL)
summary(otropais_directo$P5_DIRECTA_PERIMETRAL)

#6IND
crossp6_espanya=RRuni(response=espanya_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                    data=espanya_indirecto, model="Crosswise",p=0.37)
summary(crossp6_espanya)

crossp6_europa=RRuni(europa_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                    data=europa_indirecto, model="Crosswise",p=0.37)

```

```

summary(crossp6_europa)

crossp6_america=RRuni(response=america_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                      data=america_indirecto, model="Crosswise",p=0.37)
summary(crossp6_america)

crossp6_otropais=RRuni(response=otropais_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                      data=otropais_indirecto, model="Crosswise",p=0.37)
summary(crossp6_otropais)

#6DIR
summary(espanya_directo$P6_DIRECTA_TOQUEQUEDA)
summary(europa_directo$P6_DIRECTA_TOQUEQUEDA)
summary(america_directo$P6_DIRECTA_TOQUEQUEDA)
summary(otropais_directo$P6_DIRECTA_TOQUEQUEDA)

-----

#Situacion laboral + univariantes

trabajando_indirecto=subset(grupocrosswise, grupocrosswise$SITUAC_LAB==1)
enparo_indirecto=subset(grupocrosswise, grupocrosswise$SITUAC_LAB==2)
notrabaja_indirecto=subset(grupocrosswise, grupocrosswise$SITUAC_LAB==3)
jubilado_indirecto=subset(grupocrosswise, grupocrosswise$SITUAC_LAB==4)
otrolab_indirecto=subset(grupocrosswise, grupocrosswise$SITUAC_LAB==5)

trabajando_directo=subset(grupodirecto, grupodirecto$SITUAC_LAB==1)
enparo_directo=subset(grupodirecto, grupodirecto$SITUAC_LAB==2)
notrabaja_directo=subset(grupodirecto, grupodirecto$SITUAC_LAB==3)
jubilado_directo=subset(grupodirecto, grupodirecto$SITUAC_LAB==4)
otrolab_directo=subset(grupodirecto, grupodirecto$SITUAC_LAB==5)

#1IND
crossp1_trabajando=RRuni(response=trabajando_indirecto$P1_CROSSWISE_MASCARILLA,
                          A,
                          data=trabajando_indirecto, model="Crosswise",p=0.33)

crossp1_enparo=RRuni(response=enparo_indirecto$P1_CROSSWISE_MASCARILLA,
                     data=enparo_indirecto, model="Crosswise",p=0.33)

crossp1_notrabaja=RRuni(response=notrabaja_indirecto$P1_CROSSWISE_MASCARILLA,
                        data=notrabaja_indirecto, model="Crosswise",p=0.33)

#JUBILADOS NO HAN RESPONDIDO INDIRECTO

crossp1_otrolab=RRuni(response=otrolab_indirecto$P1_CROSSWISE_MASCARILLA,
                      data=otrolab_indirecto, model="Crosswise",p=0.33)

summary(crossp1_trabajando)
summary(crossp1_enparo)
summary(crossp1_notrabaja)

```

```
summary(crossp1_otrolab)
```

```
#1DIR
```

```
summary(trabajando_directo$P1_DIRECTA_MASCARILLA)
```

```
summary(enparo_directo$P1_DIRECTA_MASCARILLA)
```

```
summary(notrabaja_directo$P1_DIRECTA_MASCARILLA)
```

```
summary(jubilado_directo$P1_DIRECTA_MASCARILLA)
```

```
summary(otrolab_directo$P1_DIRECTA_MASCARILLA)
```

```
#2IND
```

```
crossp2_trabajando=RRuni(response=trabajando_indirecto$P2_CROSSWISE_MANOS,  
                          data=trabajando_indirecto, model="Crosswise",p=0.63)
```

```
summary(crossp2_trabajando)
```

```
crossp2_enparo=RRuni(response=enparo_indirecto$P2_CROSSWISE_MANOS,  
                      data=enparo_indirecto, model="Crosswise",p=0.63)
```

```
summary(crossp2_enparo)
```

```
crossp2_notrabaja=RRuni(response=notrabaja_indirecto$P2_CROSSWISE_MANOS,  
                         data=notrabaja_indirecto, model="Crosswise",p=0.63)
```

```
summary(crossp2_notrabaja)
```

```
#JUBILADOS NO HAN RESPONDIDO INDIRECTO
```

```
crossp2_otrolab=RRuni(response=otrolab_indirecto$P2_CROSSWISE_MANOS,  
                      data=otrolab_indirecto, model="Crosswise",p=0.63)
```

```
summary(crossp2_otrolab)
```

```
#2DIR
```

```
summary(trabajando_directo$P2_DIRECTA_MANOS)
```

```
summary(enparo_directo$P2_DIRECTA_MANOS)
```

```
summary(notrabaja_directo$P2_DIRECTA_MANOS)
```

```
summary(jubilado_directo$P2_DIRECTA_MANOS)
```

```
summary(otrolab_directo$P2_DIRECTA_MANOS)
```

```
#3IND
```

```
crossp3_trabajando=RRuni(response=trabajando_indirecto$P3_CROSSWISE_CONFINAMI  
ENTO,  
                          data=trabajando_indirecto, model="Crosswise",p=0.70)
```

```
summary(crossp3_trabajando)
```

```
crossp3_enparo=RRuni(response=enparo_indirecto$P3_CROSSWISE_CONFINAMIENTO,  
                      data=enparo_indirecto, model="Crosswise",p=0.70)
```

```
summary(crossp3_enparo)
```

```
crossp3_notrabaja=RRuni(response=notrabaja_indirecto$P3_CROSSWISE_CONFINAMIEN  
TO,  
                          data=notrabaja_indirecto, model="Crosswise",p=0.70)
```

```
summary(crossp3_notrabaja)
```



```

#JUBILADOS NO HAN RESPONDIDO INDIRECTO

crossp3_otrolab=RRuni(response=otrolab_indirecto$P3_CROSSWISE_CONFINAMIENTO,
                      data=otrolab_indirecto, model="Crosswise",p=0.70)
summary(crossp3_otrolab)

#3DIR
summary(trabajando_directo$P3_DIRECTA_CONFINAMIENTO)
summary(enparo_directo$P3_DIRECTA_CONFINAMIENTO)
summary(notrabaja_directo$P3_DIRECTA_CONFINAMIENTO)
summary(jubilado_directo$P3_DIRECTA_CONFINAMIENTO)
summary(otrolab_directo$P3_DIRECTA_CONFINAMIENTO)

#4IND
crossp4_trabajando=RRuni(response=trabajando_indirecto$P4_CROSSWISE_DISTANCIASEG,
                          data=trabajando_indirecto, model="Crosswise",p=0.57)
summary(crossp4_trabajando)

crossp4_enparo=RRuni(response=enparo_indirecto$P4_CROSSWISE_DISTANCIASEG,
                      data=enparo_indirecto, model="Crosswise",p=0.57)
summary(crossp4_enparo)

crossp4_notrabaja=RRuni(response=notrabaja_indirecto$P4_CROSSWISE_DISTANCIASEG,
                          data=notrabaja_indirecto, model="Crosswise",p=0.57)
summary(crossp4_notrabaja)

#JUBILADOS NO HAN RESPONDIDO INDIRECTO

crossp4_otrolab=RRuni(response=otrolab_indirecto$P4_CROSSWISE_DISTANCIASEG,
                      data=otrolab_indirecto, model="Crosswise",p=0.57)
summary(crossp4_otrolab)

#4DIR
summary(trabajando_directo$P4_DIRECTA_DISTANCIASEG)
summary(enparo_directo$P4_DIRECTA_DISTANCIASEG)
summary(notrabaja_directo$P4_DIRECTA_DISTANCIASEG)
summary(jubilado_directo$P4_DIRECTA_DISTANCIASEG)
summary(otrolab_directo$P4_DIRECTA_DISTANCIASEG)

#5IND
crossp5_trabajando=RRuni(response=trabajando_indirecto$P5_CROSSWISE_PERIMETRAL,
                          data=trabajando_indirecto, model="Crosswise",p=0.32)
summary(crossp5_trabajando)

crossp5_enparo=RRuni(response=enparo_indirecto$P5_CROSSWISE_PERIMETRAL,
                      data=enparo_indirecto, model="Crosswise",p=0.32)
summary(crossp5_enparo)

```

```

crossp5_notrabaja=RRuni(response=notrabaja_indirecto$P5_CROSSWISE_PERIMETRAL,
                        data=notrabaja_indirecto, model="Crosswise",p=0.32)
summary(crossp5_notrabaja)

#JUBILADOS NO HAN RESPONDIDO INDIRECTO

crossp5_otrolab=RRuni(response=otrolab_indirecto$P5_CROSSWISE_PERIMETRAL,
                      data=otrolab_indirecto, model="Crosswise",p=0.32)
summary(crossp5_otrolab)

#5DIR
summary(trabajando_directo$P5_DIRECTA_PERIMETRAL)
summary(enparo_directo$P5_DIRECTA_PERIMETRAL)
summary(notrabaja_directo$P5_DIRECTA_PERIMETRAL)
summary(jubilado_directo$P5_DIRECTA_PERIMETRAL)
summary(otrolab_directo$P5_DIRECTA_PERIMETRAL)

#6IND
crossp6_trabajando=RRuni(response=trabajando_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                          data=trabajando_indirecto, model="Crosswise",p=0.37)
summary(crossp6_trabajando)

crossp6_enparo=RRuni(response=enparo_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                     data=enparo_indirecto, model="Crosswise",p=0.37)
summary(crossp6_enparo)

crossp6_notrabaja=RRuni(response=notrabaja_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                        data=notrabaja_indirecto, model="Crosswise",p=0.37)
summary(crossp6_notrabaja)

#JUBILADOS NO HAN RESPONDIDO INDIRECTO

crossp6_otrolab=RRuni(response=otrolab_indirecto$P6_CROSSWISE_TOQUEQUEDA,
                      data=otrolab_indirecto, model="Crosswise",p=0.37)
summary(crossp6_otrolab)

#6DIR
summary(trabajando_directo$P6_DIRECTA_TOQUEQUEDA)
summary(enparo_directo$P6_DIRECTA_TOQUEQUEDA)
summary(notrabaja_directo$P6_DIRECTA_TOQUEQUEDA)
summary(jubilado_directo$P6_DIRECTA_TOQUEQUEDA)
summary(otrolab_directo$P6_DIRECTA_TOQUEQUEDA)
-----

#regresion

#modelo_p1=RRlog(P1_CROSSWISE_MASCARILLA~SEXO+EDAD+NACIONALIDAD,
data=grupocrosswise,
                #model="Crosswise",p=0.33)

#modelo_p2=RRlog(P2_CROSSWISE_MANOS~SEXO+EDAD+NACIONALIDAD,
data=grupocrosswise,

```

```

#model="Crosswise",p=0.63)

#modelo_p3=RRlog(P3_CROSSWISE_CONFINAMIENTO~SEXO+EDAD+NACIONALIDAD,
data=grupocrosswise,
#model="Crosswise",p=0.70)

#modelo_p4=RRlog(P4_CROSSWISE_DISTANCIASEG~SEXO+EDAD+NACIONALIDAD,
data=grupocrosswise,
#model="Crosswise",p=0.57)

#modelo_p5=RRlog(P5_CROSSWISE_PERIMETRAL~SEXO+EDAD+NACIONALIDAD,
data=grupocrosswise,
#model="Crosswise",p=0.32)

#modelo_p6=RRlog(P6_CROSSWISE_TOQUEQUEDA~SEXO+EDAD+NACIONALIDAD,
data=grupocrosswise,
#model="Crosswise",p=0.63)

#contrastes sexo (directas)

#medias y varianzas

n0_mujeresdirecto=93
n1_mujeresindirecto=119
n0_hombresdirecto=92
n1_hombresindirecto=134
Nhombre=24679
Nmujer=39943
pi_sexo_directo=rep(n0_mujeresdirecto/Nhombre,n0_hombresdirecto)
pi_sexo_indirecto=rep(n1_mujeresindirecto/Nmujer,n1_hombresindirecto)

#1 DIRECTO
HTP1DIRECTA_mujdir<-HTestimator(Mujeres_directo$P1_DIRECTA_MASCARILLA,pik =
pi_sexo_directo)/Nmujer

VARP1DIRECTA_1_mujdir<-varest(Mujeres_directo$P1_DIRECTA_MASCARILLA,pik =
pi_sexo_directo)/Nmujer^2

HTP1DIRECTA_homdir<-HTestimator(Hombres_directo$P1_DIRECTA_MASCARILLA,pik =
pi_sexo_directo)/Nhombre

VARP1DIRECTA_1_homdir<-varest(Hombres_directo$P1_DIRECTA_MASCARILLA,pik =
pi_sexo_directo)/Nhombre^2

#2 DIRECTO
HTP2DIRECTA_mujdir<-HTestimator(Mujeres_directo$P2_DIRECTA_MANOS,pik =
pi_sexo_directo)/Nmujer

VARP2DIRECTA_2_mujdir<-varest(Mujeres_directo$P2_DIRECTA_MANOS,pik =
pi_sexo_directo)/Nmujer^2

```

```

HTP2DIRECTA_homdir<-HTestimator(Hombres_directo$P2_DIRECTA_MANOS,pik      =
pi_sexo_directo)/Nhombre

VARP2DIRECTA_2_homdir<-varest(Hombres_directo$P2_DIRECTA_MANOS,pik      =
pi_sexo_directo)/Nhombre^2

#3
HTP3DIRECTA_mujdir<-HTestimator(Mujeres_directo$P3_DIRECTA_CONFINAMIENTO,pik
= pi_sexo_directo)/Nmujer

VARP3DIRECTA_3_mujdir<-varest(Mujeres_directo$P3_DIRECTA_CONFINAMIENTO,pik  =
pi_sexo_directo)/Nmujer^2

HTP3DIRECTA_homdir<-HTestimator(Hombres_directo$P3_DIRECTA_CONFINAMIENTO,pik
= pi_sexo_directo)/Nhombre

VARP3DIRECTA_3_homdir<-varest(Hombres_directo$P3_DIRECTA_CONFINAMIENTO,pik  =
pi_sexo_directo)/Nhombre^2

#4
HTP4DIRECTA_mujdir<-HTestimator(Mujeres_directo$P4_DIRECTA_DISTANCIASEG,pik =
pi_sexo_directo)/Nmujer

VARP4DIRECTA_4_mujdir<-varest(Mujeres_directo$P4_DIRECTA_DISTANCIASEG,pik  =
pi_sexo_directo)/Nmujer^2

HTP4DIRECTA_homdir<-HTestimator(Hombres_directo$P4_DIRECTA_DISTANCIASEG,pik =
pi_sexo_directo)/Nhombre

VARP4DIRECTA_4_homdir<-varest(Hombres_directo$P4_DIRECTA_DISTANCIASEG,pik  =
pi_sexo_directo)/Nhombre^2

#5
HTP5DIRECTA_mujdir<-HTestimator(Mujeres_directo$P5_DIRECTA_PERIMETRAL,pik  =
pi_sexo_directo)/Nmujer

VARP5DIRECTA_5_mujdir<-varest(Mujeres_directo$P5_DIRECTA_PERIMETRAL,pik  =
pi_sexo_directo)/Nmujer^2

HTP5DIRECTA_homdir<-HTestimator(Hombres_directo$P5_DIRECTA_PERIMETRAL,pik  =
pi_sexo_directo)/Nhombre

VARP5DIRECTA_5_homdir<-varest(Hombres_directo$P5_DIRECTA_PERIMETRAL,pik  =
pi_sexo_directo)/Nhombre^2

#6
HTP6DIRECTA_mujdir<-HTestimator(Mujeres_directo$P6_DIRECTA_PERIMETRAL,pik  =
pi_sexo_directo)/Nmujer

```

```
VARP6DIRECTA_6_mujdir<-varest (Mujeres_directo$P6_DIRECTA_PERIMETRAL,pik      =  
pi_sexo_directo) /Nmujer^2
```

```
HTP6DIRECTA_homdir<-HTestimator(Hombres_directo$P6_DIRECTA_PERIMETRAL,pik    =  
pi_sexo_directo) /Nhombre
```

```
VARP6DIRECTA_6_homdir<-varest (Hombres_directo$P6_DIRECTA_PERIMETRAL,pik    =  
pi_sexo_directo) /Nhombre^2
```

```
HTP1DIRECTA_mujdir  
HTP1DIRECTA_homdir  
HTP2DIRECTA_mujdir  
HTP2DIRECTA_homdir  
HTP3DIRECTA_mujdir  
HTP3DIRECTA_homdir  
HTP4DIRECTA_mujdir  
HTP4DIRECTA_homdir  
HTP5DIRECTA_mujdir  
HTP5DIRECTA_homdir  
HTP6DIRECTA_mujdir  
HTP6DIRECTA_homdir  
VARP1DIRECTA_1_mujdir  
VARP2DIRECTA_2_mujdir  
VARP3DIRECTA_3_mujdir  
VARP4DIRECTA_4_mujdir  
VARP5DIRECTA_5_mujdir  
VARP6DIRECTA_6_mujdir  
VARP1DIRECTA_1_homdir  
VARP2DIRECTA_2_homdir  
VARP3DIRECTA_3_homdir  
VARP4DIRECTA_4_homdir  
VARP5DIRECTA_5_homdir  
VARP6DIRECTA_6_homdir
```

```
#1mujer  
#1hombre
```

```
Z_p1muj<-(HTP1DIRECTA_mujdir-mujerdirecto$pi_sexo_directo)/sqrt (VARP1DIRECTA_  
1_mujdir+(mujerdirecto$pi_sexo_directo)^2)  
Z_p1hom<-(HTP1DIRECTA_homdir-hombredirecto$pi_sexo_directo)/sqrt (VARP1DIRECTA_  
1_homdir+(hombredirecto$pi_sexo_directo)^2)
```

```
#2mujer  
#2hombre
```

```
Z_p2muj<-(HTP2DIRECTA_mujdir-mujerdirecto$pi_sexo_directo)/sqrt (VARP2DIRECTA_  
2_mujdir+(mujerdirecto$pi_sexo_directo)^2)  
Z_p2hom<-(HTP2DIRECTA_homdir-hombredirecto$pi_sexo_directo)/sqrt (VARP2DIRECTA_  
2_homdir+(hombredirecto$pi_sexo_directo)^2)
```

```

#3hombre
#3mujer

Z_p3muj<- (HTP3DIRECTA_mujdir-mujerdirecto$pi_sexo_directo)/sqrt (VARP3DIRECTA_
3_mujdir+(mujerdirecto$pi_sexo_directo)^2)
Z_p3hom<- (HTP3DIRECTA_homdir-hombredirecto$pi_sexo_directo)/sqrt (VARP3DIRECTA
_3_homdir+(hombredirecto$pi_sexo_directo)^2)

#4hombre
#4mujer

Z_p4muj<- (HTP4DIRECTA_mujdir-mujerdirecto$pi_sexo_directo)/sqrt (VARP4DIRECTA_
4_mujdir+(mujerdirecto$pi_sexo_directo)^2)
Z_p4hom<- (HTP4DIRECTA_homdir-hombredirecto$pi_sexo_directo)/sqrt (VARP4DIRECTA
_4_homdir+(hombredirecto$pi_sexo_directo)^2)

#5hombre
#5mujer

Z_p5muj<- (HTP5DIRECTA_mujdir-mujerdirecto$pi_sexo_directo)/sqrt (VARP5DIRECTA_
5_mujdir+(mujerdirecto$pi_sexo_directo)^2)
Z_p5hom<- (HTP5DIRECTA_homdir-hombredirecto$pi_sexo_directo)/sqrt (VARP5DIRECTA
_5_homdir+(hombredirecto$pi_sexo_directo)^2)

#6hombre
#6mujer

Z_p6muj<- (HTP6DIRECTA_mujdir-mujerdirecto$pi_sexo_directo)/sqrt (VARP6DIRECTA_
6_mujdir+(mujerdirecto$pi_sexo_directo)^2)
Z_p6hom<- (HTP6DIRECTA_homdir-hombredirecto$pi_sexo_directo)/sqrt (VARP6DIRECTA
_6_homdir+(hombredirecto$pi_sexo_directo)^2)

#P valor

#1
pvalorp1_hom<-2*pnorm (abs (Z_p1hom) , lower.tail=FALSE)
pvalorp1_muj<-2*pnorm (abs (Z_p1muj) , lower.tail=FALSE)

#2
pvalorp2_hom<-2*pnorm (abs (Z_p2hom) , lower.tail=FALSE)
pvalorp2_muj<-2*pnorm (abs (Z_p2muj) , lower.tail=FALSE)

#3
pvalorp3_hom<-2*pnorm (abs (Z_p3hom) , lower.tail=FALSE)
pvalorp3_muj<-2*pnorm (abs (Z_p3muj) , lower.tail=FALSE)

#4

```

```
pvalorp4_hom<-2*pnorm(abs(Z_p4hom),lower.tail=FALSE)
pvalorp4_muj<-2*pnorm(abs(Z_p4muj),lower.tail=FALSE)
```

```
#5
```

```
pvalorp5_hom<-2*pnorm(abs(Z_p5hom),lower.tail=FALSE)
pvalorp5_muj<-2*pnorm(abs(Z_p5muj),lower.tail=FALSE)
```

```
#6
```

```
pvalorp6_hom<-2*pnorm(abs(Z_p6hom),lower.tail=FALSE)
pvalorp6_muj<-2*pnorm(abs(Z_p6muj),lower.tail=FALSE)
```