



# MÁSTER DE ESTADÍSTICA APLICADA CON R SOFTWARE

## RESULTADO DE UN EXPERIMENTO PARA ANALIZAR LA INFLUENCIA DEL FORMATO DE LA ESCALA DE RESPUESTA EN LA MEDICIÓN DE ACTITUDES EN ENCUESTAS DE OPINIÓN

AUTORA: Ana Muñoz van den Eynde

DIRECTOR: Juan Luis López Garrancho

FECHA: 24 de enero de 2020

ENTIDAD COLABORADORA: MÁXIMA Formación

## **RESUMEN**

Las encuestas de opinión son una herramienta importante para saber qué piensa, siente o sabe el público sobre un tema, por lo que garantizar la más alta calidad de la información que brindan es crucial. No obstante, se ha encontrado que las respuestas proporcionadas dependen mucho del contexto de la encuesta, es decir, del orden de las preguntas y de la forma en que se formulan, como resultado de que la situación de responder a una encuesta parece poner en marcha el sistema de procesamiento de la información intuitivo, rápido y en paralelo. En este contexto, este trabajo plantea dos objetivos: 1) analizar si el formato de la escala de respuesta genera estimaciones diferentes de la intensidad de las actitudes; 2) analizar si el formato de respuesta influye en la capacidad de discriminación de las preguntas diseñadas para medir actitudes. Para alcanzar este objetivo se diseñó un experimento con 8 ítems divididos en dos preguntas y 8 formatos de respuesta diferentes que se distribuyeron aleatoriamente entre 873 voluntarios. Los resultados del experimento realizado indican que la calidad de las respuestas, entendida como mayor precisión, aumenta cuando lo hace el número de opciones de la escala de respuesta y la dificultad de la pregunta. Por tanto, para contribuir a que la actividad de responder a una encuesta esté dirigida por el sistema 2 de procesamiento, debemos generar opciones de respuesta que faciliten que ese sistema tome el control sobre el automático, rápido y asociativo.

## ÍNDICE

Resumen .....	1
1 INTRODUCCIÓN .....	3
1.1 Las encuestas de opinión.....	3
1.2 Actitudes.....	5
1.3 El proceso cognitivo de responder a una encuesta .....	7
1.4 Escalas de respuesta de Acuerdo/Desacuerdo .....	7
1.5 Las opciones de respuesta .....	8
A) El número de categorías.....	8
B) La presencia de un punto intermedio neutral .....	10
C) Las etiquetas de las opciones de respuesta .....	11
1.6 Estilos de respuesta .....	11
1.7 Algunas consideraciones metodológicas .....	13
1.8 Objetivo .....	15
2 DESARROLLO DEL TRABAJO .....	15
2.1 Diseño del experimento.....	15
2.2 Ajuste del fichero de datos .....	17
2.3 Análisis estadístico.....	18
2.4 Hipótesis.....	19
2.5 Procedimiento .....	19
2.6 Resultados .....	20
A) Variable dependiente: suma de las puntuaciones transformadas (sumT).....	20
B) Variable dependiente: variabilidad.....	23
C) Análisis de las diferencias entre las preguntas 1 y 2.....	26
3 CONCLUSIONES.....	29
BIBLIOGRAFÍA .....	32
ANEXO I – Diseño del experimento .....	36
ANEXO II – Código para el ajuste del fichero de datos .....	53
ANEXO III – Análisis de la variable sumT en la pregunta 1 .....	57
ANEXO IV - Análisis de la variable sumT en la pregunta 2 .....	66
ANEXO V – Análisis de la variabilidad en la pregunta1 .....	75
ANEXO VI – Análisis de la variabilidad en la pregunta2.....	83
ANEXO VII – Análisis de las diferencias entre las preguntas 1 y 2 .....	91

## 1. INTRODUCCIÓN

### 1.1 Las encuestas de opinión

Las encuestas de opinión son una herramienta importante para saber qué piensa, siente o sabe el público sobre un tema. Garantizar la más alta calidad de la información que brindan es crucial. No obstante, se ha encontrado que las respuestas que las personas proporcionan al responder una encuesta dependen mucho del contexto de esta, es decir, del orden de las preguntas y de la forma en que se formulan. Estos efectos no ocurren al azar sino sistemáticamente (Tourangeau et al., 2000). También se ha descubierto que las personas son capaces de crear una respuesta cuando se les pregunta sobre problemas ficticios. En estos contextos experimentales, las personas que tratan de ofrecer respuestas de calidad parecen estar involucradas en una búsqueda activa de significado que extraen tanto de la redacción de la pregunta como del contexto general en el que se formula (Sturgis y Smith, 2010).

Cada uno de nosotros está sometido de forma constante a grandes cantidades de información de muchos tipos. Se llama cognición al conjunto de actividades a través de las cuales se procesa esta información. Este proceso no depende totalmente de las características *objetivas*, es más bien la reconstrucción mental de lo que es real realizada por los individuos a partir de su experiencia pasada, sus necesidades, deseos e intenciones (Leyens y Codol, 1992). Desde la psicología cognitiva se señala que los seres humanos disponemos de dos sistemas para procesar la información. El sistema 1 es rápido, automático e implica poco o ningún esfuerzo cognitivo. Genera impresiones y sentimientos. Y, en la búsqueda de la eficiencia cognitiva, actúa por defecto. Además, no puede ser desconectado. El sistema 2 está bajo control cognitivo, es el que distribuye la atención y, por tanto, se ocupa de las actividades mentales que requieren esfuerzo utilizando como *materia prima* las impresiones y sentimientos generados por el sistema 1 (Kahneman, 2011). En los estudios mediante encuestas se ha partido del supuesto de que responder a una encuesta es una actividad dirigida por el sistema 2, sin embargo, cada vez hay más evidencia de que la situación de responder a una encuesta pone en marcha el sistema de procesamiento de la información intuitivo, rápido y en paralelo. Por ese motivo, el enunciado de la pregunta es fundamental, ya que condiciona el procesamiento de la información incluida en él y, por tanto, la respuesta que se obtiene (Tourangeau et al., 2000). Por ejemplo, se ha encontrado que el simple hecho de preguntar por un tema concreto hace que este esté más presente y, por tanto, se piense en él con más dedicación y seriedad de lo que

hacemos en nuestro día a día. Por otro lado, las personas con posiciones más extremas o definidas son menos susceptibles a estos efectos. En cambio, las personas para las que un tema concreto es poco importante, son especialmente sensibles a la influencia del contexto (Van der Pligt et al., 1987).

La rendición de cuentas (*accountability*, en inglés) es la expectativa, implícita o explícita, de que debemos justificar nuestras creencias, sentimientos y acciones ante los demás. Este efecto influye de manera fundamental en los procesos cognitivos implicados en la percepción, codificación y recuperación de información. Es posible que no influya en cómo piensan las personas, pero sí en lo que dicen que piensan; hasta el punto de que pueden cambiar su opinión, al menos temporalmente (Lerner y Tetlock, 1999). Y pueden *crear* una opinión cuando se les pregunta, porque es lo que se espera de ellos cuando responden a una encuesta. O una explicación para justificar la opinión manifestada, cuando en realidad su respuesta es resultado del procesamiento asociativo. En este sentido, se ha encontrado que las respuestas a las encuestas no son el producto de prolongadas deliberaciones; al contrario, las personas encuestadas tardan menos de cinco segundos en responder a las típicas cuestiones de opinión (Tourangeau et al., 2000).

Por último, hay una ley general del “mínimo esfuerzo” que guía tanto la actividad física como la cognitiva. De acuerdo con ella, si hay varias opciones para lograr un mismo objetivo, la persona elegirá la pauta de acción menos exigente (Kahneman, 2011). Se ha acuñado el adjetivo *satisficiente*, combinación de satisfactorio y suficiente, para describir la estrategia de optimizar recursos cognitivos (Krosnick y Alwin, 1987; Krosnick, 1991). De acuerdo con esta estrategia, al responder a una encuesta no se invierten los recursos necesarios para proporcionar respuestas óptimas. Al contrario, se buscan pistas en las preguntas que señalen respuestas razonables que seleccionar rápidamente. Evidentemente, todos estos efectos son más relevantes cuanto menor sea el esfuerzo cognitivo que podemos solicitarle a las personas encuestadas. Como han demostrado multitud de trabajos sobre cognición social, las estrategias utilizadas a la hora de formular juicios sobre opiniones dependen de la motivación que asignen los sujetos a la tarea. Cuanto más implicados se sientan hacia ella, mayor es la probabilidad de que adopten una estrategia sistemática de procesamiento (mayor control por parte del sistema 2). Por el contrario, cuando las personas se sienten poco implicadas, mayor es la probabilidad de utilizar estrategias de procesamiento heurísticas, prestando toda la atención a la información que viene rápidamente a la mente (Schwarz y Vaughn, 2002).

Como resultado de estos procesos, las opiniones manifestadas están sujetas a una variedad de *efectos de respuesta* bien documentados. Como han señalado Tourangeau et al. (2000), se trata de diferencias en las respuestas que no tienen que ver con la opinión de quien responde, sino que son consecuencia del propio proceso de responder a la encuesta y del contexto. Se pueden deber a dificultades para entender la pregunta o recordar la información necesaria, a la capacidad para proporcionar una respuesta adecuada, u a otros procesos cognitivos. Aunque el término efectos de respuesta puede indicar que son voluntarios, los autores consideran que son el resultado de las estrategias automáticas del procesamiento cognitivo. Así mismo, identifican cuatro etapas en el proceso de responder a una encuesta que son susceptibles de verse influidas por estos efectos: comprensión, recuperación, juicio y respuesta. La especificación de estas fases no significa que quien responde pase necesariamente por todas ellas. Tanto los procesos que se llevan a cabo como los resultados generados dependen de multitud de factores, entre los que se pueden mencionar la voluntad de ser precisos en las respuestas, o la cantidad de tiempo de que se disponga para responder (Muñoz van den Eynde, 2014).

## **1.2 Actitudes**

Las encuestas de opinión dedican mucho esfuerzo a medir las actitudes de quienes responden en relación con multitud de temas. Sin embargo, a pesar de ser un elemento constante en las encuestas de opinión, su conceptualización y medición presenta importantes dificultades. Las investigaciones sobre las actitudes han asumido que las personas tienen una actitud formada sobre todas las cuestiones por las que se les pregunta, es decir, tienen almacenada en la memoria una respuesta para las preguntas de las encuestas de opinión y solo necesitan recuperarla cuando se les pide. Sin embargo, los resultados proporcionados por la literatura sobre el tema presentan una imagen bastante diferente. Lo más llamativo ha sido la confirmada influencia del contexto en las respuestas y, más en concreto, la influencia del orden de las preguntas (Tourangeau et al., 2000). Si para responder a una pregunta solo hay que recuperar una opinión almacenada en la memoria, el orden en que se presentan no tendría que ser determinante. Se considera que ese es uno de los principales problemas del estudio de las actitudes. De hecho, los últimos desarrollos en la investigación sobre el tema tienen cada vez más en cuenta estas cuestiones, hasta el punto de que se ha llegado a cuestionar su existencia. Se considera también poco probable que haya, tal y como se ha postulado, actitudes específicas para todo tipo de objetos de actitud. Los resultados mencionados indican que es más lógico pensar en la existencia de actitudes generales entendidas como predisposiciones, es decir, en el

sentido, recogido en el Diccionario de la Real Academia Española de la Lengua, de preparar o disponer de manera anticipada algo o el ánimo de alguien para un fin determinado (Muñoz van den Eynde, 2014). Por otro lado, la mayor parte de las teorías sobre las actitudes las conciben como constructos bipolares, con un polo positivo y otro negativo, aunque algunos autores hablan también de la existencia de actitudes unipolares, de manera que hay un polo neutral y otro positivo o negativo, pero no ambos (Pratkanis y Greenwald, 1989).

Con independencia de lo anterior, la forma más utilizada para medir las actitudes es una escala de respuesta bipolar en la que se pide a los encuestados que califiquen hasta qué punto están de acuerdo o en desacuerdo con una frase supuestamente orientada a capturar la orientación positiva o negativa de quien responde hacia el objeto de actitud (Likert, 1932). Para la construcción de estas escalas es fundamental elegir si se usa un número par o impar de alternativas de respuesta. Si se selecciona un número impar, se debe suponer que el punto medio de la escala de respuesta representa puntos de vista que son realmente neutrales (Sturgis, Roberts y Smith, 2014). Sin embargo, se ha encontrado que los encuestados tienden a seleccionar la alternativa de respuesta de punto medio para manifestar incertidumbre sin tener que admitir ignorancia (Sturgis y Smith, 2010). También hay evidencia de que las personas tienden a seleccionar las respuestas "ni de acuerdo / ni en desacuerdo" (NA / ND) para reflejar la falta de opinión en lugar de la neutralidad de opinión (Sturgis, Roberts y Smith, 2012). La tendencia a seleccionar la opción que en realidad no manifiesta una opinión es mayor entre los encuestados con menos conocimiento sobre el tema sobre el que se les pregunta (Krosnick et al., 2002). Parece también que este tipo de respuestas están relacionadas con falta de información o interés sobre el tema de la encuesta (Allum et al., 2008).

En todo caso, los encuestados que no tienen una actitud u opinión firme sobre un tema concreto, ya sea porque tienen poco conocimiento del tema o porque no hayan pensado en él previamente, pueden responder de manera aleatoria cuando se enfrentan a la consiguiente pregunta de actitud. Por otro lado, y según la teoría del juicio social, las actitudes no se ven representadas por un punto concreto en el continuo de actitud; más bien, las personas tienen "latitudes de aceptación". La latitud de aceptación constituye la región de una escala (o conjunto de categorías de respuesta) dentro de la cual un encuestado ubicaría presumiblemente su actitud sobre el objeto por el que se le pregunta. Cuanto mayor sea la latitud de aceptación, mayor será el rango dentro del cual el encuestado podría responder; obligarles a elegir un solo punto, que presumiblemente tendría un alto componente de imprecisión, puede

llevar a los encuestados a tomar su decisión al azar, aumentando la cantidad de error aleatorio en la medición. Por otro lado, es posible que los encuestados tengan dificultades para elegir un solo punto en la escala de respuesta incluida en la pregunta. Algunos errores de respuesta pueden, por lo tanto, estar asociados con juicios ambiguos o conflictivos con respecto al vínculo entre las señales internas de quien responde y las categorías de respuesta proporcionadas externamente (Alwin y Krosnick, 1991).

### **1.3 El proceso cognitivo de responder a una encuesta**

En el proceso de proporcionar una respuesta significativa, los encuestados deben inferir las intenciones del interlocutor, es decir, el significado pragmático de la pregunta (Schwarz 2007). De acuerdo con el *Modelo del Proceso de Respuesta*, al responder una pregunta los encuestados deben: (a) comprender la pregunta, (b) recuperar información relevante de la memoria a largo plazo, (c) emitir un juicio basado en la información recuperada y (d) ajustar su respuesta a las opciones proporcionadas por quien ha diseñado la encuesta. Estos cuatro procesos son potencialmente problemáticos para cualquier encuestado (Tourangeau et al., 2000). De hecho, numerosos estudios han demostrado que las personas encuestadas se basan en aspectos "formales" del diseño del cuestionario, desde la elección de alternativas de respuesta hasta las características formales de las escalas y el diseño gráfico de los cuestionarios, para proporcionar su respuesta (Schwarz, 2007).

Por otro lado, de acuerdo con Searle (1969), hacer una pregunta implica la asunción de ciertas "condiciones de sinceridad" por parte de los participantes, una de las cuales es la suposición de que la respuesta a la pregunta debe estar entre las opciones de respuesta proporcionadas por el investigador. En este sentido, se ha encontrado de manera consistente en la literatura que los encuestados generalmente actúan de acuerdo con el supuesto de esta condición de sinceridad y, en consecuencia tienden a ser reacios a ofrecer una "no opinión" a menos que se incluya explícitamente en las opciones de respuesta (Tourangeau et al., 2000). Esta también puede ser la explicación del hallazgo de que es más probable que los encuestados elijan la categoría intermedia cuando se ofrece explícitamente (por ejemplo, Schuman y Presser, 1996).

### **1.4 Escalas de respuesta Acuerdo/ Desacuerdo (A/D)**

Las escalas A/D se usan ampliamente en los cuestionarios de investigación en ciencias sociales a pesar del hecho de que las respuestas se ven afectadas por

diversos sesgos y limitaciones (Revilla et al., 2013). Por ejemplo, son susceptibles al sesgo de respuesta de aquiescencia, es decir, la tendencia a manifestarse de acuerdo con la pregunta con independencia de su contenido (p. e. Krosnick, 1991). Por otro lado, responder a este tipo de preguntas requiere un procesamiento cognitivo aún más complejo del que ya se deriva de la propia tarea de responder una encuesta (Revilla et al., 2013). En las escalas A/D, los encuestados deben identificar qué es exactamente lo que les están preguntando para seleccionar después cuál de las opciones de respuesta ofrecidas es la que mejor les permite expresar su opinión sobre esa cuestión. Como consecuencia de los grados de libertad proporcionados, cada persona puede hacer una elección diferente teniendo, en cambio, una misma opinión (Revilla et al., 2013; Saris et al. 2010). En consecuencia, se ha encontrado que cuando se les pregunta sobre las razones detrás de sus respuestas a las escalas A/D, algunos encuestados pueden proporcionar la misma razón para justificar respuestas opuestas o usar la misma opción de respuesta para expresar opiniones opuestas (Hipkins et al., 2002). Además, se ha encontrado que los encuestados tienden a desarrollar patrones de respuesta ante este tipo de preguntas que varían de un individuo a otro (Saris et al., 2010).

A pesar de estos problemas, las escalas A/D se usan ampliamente por razones prácticas, ya que la misma escala se puede usar para medir una amplia gama de constructos teóricos. Una vez adoptado el formato de escala A/D, hay otras decisiones que tomar: el número de opciones, etiquetarlas o no etiquetarlas, que la escala sea bipolar o unipolar, incluir o no una opción intermedia neutra, o cómo definir los extremos son cuestiones importantes a tener en cuenta, pues hay abundante evidencia de que estas decisiones influyen en las respuestas que se obtienen.

### **1.5 Las opciones de respuesta**

Aunque está ampliamente documentado que pequeñas variaciones en el formato de la pregunta pueden producir cambios importantes en las respuestas, la evidencia no termina de ser concluyente y, por tanto, sigue sin haber consenso sobre cuál es el mejor formato de respuesta.

#### **A) El número de categorías**

Se reconoce cada vez más que los errores de medición en cualquier tipo de datos (incluidos los datos de investigación mediante encuestas) pueden tener profundos efectos en las relaciones identificadas estadísticamente. El número de categorías de escala de respuesta parece tener importantes efectos en la calidad de los datos; y, en

general, la evidencia apunta a que la calidad aumenta al hacerlo el número de categorías de respuesta (Andrews, 1984). Aunque la realidad es bastante compleja y, por eso, a pesar de décadas de investigación, la identificación del número óptimo de categorías de respuesta sigue sin haber quedado establecido (Preston y Colman, 2000).

Los teóricos de la información han argumentado que cuantas más opciones de respuesta, mejor, ya que más categorías pueden transmitir más bits de información. Los teóricos cognitivos, por otro lado, sugieren que hay un límite superior práctico para el número de categorías, dadas las dificultades potenciales que los encuestados pueden tener para discriminar entre un mayor número de opciones y para seleccionar aquella que refleje adecuadamente su actitud. Además, los teóricos de la motivación han argumentado que los encuestados pueden no estar interesados en hacer distinciones óptimas en la calificación de los objetos de actitud o en hacer juicios o preferencias. Se ha sugerido que muchos encuestados pueden tender a proporcionar una respuesta satisfactoria (producida por el sistema 1 de procesamiento) en lugar de una respuesta óptima (generada por el sistema 2) (Alwin, 1992).

En términos generales, existe bastante consenso en considerar que la escala no debe estar tan escasamente definida (número reducido de categorías) como para perder parte de la capacidad que se atribuye a las personas encuestadas, ni ser tan fina que se agregue error de medida a los datos obtenidos si las categorías de escala requieren una precisión que excede la capacidad de los encuestados de diferenciar en sus opiniones; en este sentido, hay evidencia que sugiere que la cantidad de información transmitida por una escala de respuesta aumenta al hacerlo el número de categorías, pero este incremento se va haciendo cada vez más pequeño (Bendig, 1953). Por tanto, parece factible considerar que hay un límite para el beneficio de agregar categorías de respuesta o puntos de escala. Debido a que las personas probablemente diferencian entre sentimientos débiles, moderados y fuertes hacia los objetos de actitud, las escalas de respuesta de 7 puntos parecen preferibles a las más cortas. Sin embargo, una vez que crecen por encima de esta cifra, parece probable que el significado específico de los puntos de la escala gane en ambigüedad de manera progresiva (Alwin y Krosnick, 1991). Tal y como han recordado Preston y Colman (2000), hay evidencia de que la memoria a corto plazo de los seres humanos tiene capacidad para procesar siete elementos diferentes (más o menos dos), lo que implica un límite de aproximadamente siete en el número de categorías que las personas pueden usar para emitir juicios sobre la magnitud de estímulos unidimensionales. Esto tiene implicaciones para las escalas de respuesta, en el

sentido de que indicaría que no se obtiene información adicional al aumentar el número de categorías de respuesta por encima de siete.

Alwin y Krosnick (1991) encontraron que las preguntas que utilizaban escalas A/D con más opciones de respuesta tendían a mostrar mayor fiabilidad, aunque la realidad era algo más compleja, pues había una interacción con el etiquetado de las opciones de respuesta, de tal manera que utilizar etiquetas para todas las categorías de la escala se asociaba con una mayor fiabilidad que etiquetar solo los extremos. También encontraron que algunas características de los encuestados se relacionaban con los niveles de fiabilidad; en concreto, observaron que las respuestas de los encuestados de más edad y/o con menos nivel de estudios eran menos fiables. También encontraron que los resultados que se obtienen al medir la dirección de las actitudes son más fiables que los obtenidos al medir su intensidad.

Preston y Colman (2000) encontraron que la fiabilidad de las escalas es máxima cuando tienen siete categorías. Por otro lado, estos autores también midieron las preferencias de los encuestados (una cuestión insuficientemente explorada en la literatura) en tres dimensiones: facilidad de uso, rapidez y capacidad para expresar la opinión de manera adecuada y encontraron que la escala que mejor resultado obtuvo teniendo en cuenta estas tres dimensiones fue la de 10 puntos, seguida de cerca por la de 7 y 9 puntos. Por tanto, concluyeron que el número óptimo de categorías en una escala de opinión se sitúa entre 7, 9 y 10.

### **B) La presencia de un punto intermedio neutral**

Una cuestión importante a la hora de definir el formato de la escala de respuesta es la conveniencia de incluir una respuesta intermedia neutral. Por un lado, esto a veces se considera desaconsejable porque proporciona un escape demasiado fácil y atractivo para los encuestados que no están dispuestos a expresar una opinión definitiva (Mattel y Jacoby, 1972). Por otro lado, forzar las respuestas a un formato de acuerdo o desacuerdo puede causar dificultades para muchos encuestados (Mattel y Jacoby, 1972). Es probable que los encuestados que no tienen una actitud formada hacia un objeto o que tienen sentimientos genuinamente ambivalentes prefieran colocarse en el punto medio del continuo evaluativo. Sin embargo, si se enfrentan a una escala de respuesta con un número par de alternativas de respuesta, se verían obligados a elegir entre representar su opinión como débilmente favorable o débilmente desfavorable hacia el objeto de actitud. Esta opción a menudo puede ser aleatoria, por lo que ofrecer una alternativa intermedia a los encuestados probablemente aumentaría

la fiabilidad de las respuestas, aunque no hay duda de que el punto medio se elegiría con más frecuencia (Alwin y Krosnick, 1991).

La evidencia sugiere que algunas personas que seleccionan el punto neutral en una escala no necesariamente se consideran neutrales hacia el objeto de actitud. Una posible explicación podría ser la tosquedad del instrumento empleado. Por ejemplo, el encuestado puede encontrar que no tiene una actitud suficientemente positiva o negativa hacia un objeto para respaldar la opinión positiva o negativa más baja en el continuo de una escala gruesa. Sin embargo, si se proporciona una escala más fina, permitiendo así que el sujeto exprese sus actitudes con mayor precisión, la utilización del punto neutral puede disminuir notablemente. Este es el resultado que encontraron Mattel y Jacoby (1972).

### **C) Las etiquetas de las opciones de respuesta**

Parece lógico asumir que la numeración de las escalas de respuesta lleva inherente cierta ambigüedad en su significado, es decir, ¿cómo podemos estar seguros de que el valor “5” significa lo mismo para personas distintas? Incorporar algún tipo de etiqueta verbal a las opciones de respuesta numéricas, que contribuya a definir las de alguna manera, probablemente contribuya a aclarar el significado de las alternativas ofrecidas. No obstante, es también probable que haya cierta ambigüedad inherente incluso en las definiciones verbales. Varios estudios respaldan la hipótesis de que aumentar la proporción de los puntos de escala que se etiquetan verbalmente aumenta la fiabilidad del ítem (Alwin y Krosnick, 1991). En este sentido, Bendig (1953) encontró que aumentar el etiquetado verbal de las categorías aumenta la fiabilidad y la cantidad de información transmitida por la escala. No obstante, se ha encontrado que el tipo de etiqueta utilizado en los extremos de la escala también influye en los resultados, de manera que cuando estos se expresan en términos más absolutos (totalmente versus muy) se observa una concentración de respuestas en los valores intermedios (Leung, 2011).

### **1.6 Estilos de respuesta**

Tourangeau et al. (2000) han identificado ciertos rasgos que tienden a repetirse en las respuestas proporcionadas por las personas encuestadas. El sesgo de positividad hace referencia a la tendencia a elegir el extremo positivo de una escala bipolar. La influencia del rango de las etiquetas numéricas indica que es más probable que las respuestas se acumulen en la mitad positiva de la escala bipolar (en línea con el sesgo de positividad) si estas incluyen un segmento negativo y otro positivo claramente

identificado (p. e. de -5 a +5), que si van de 0 a 10 aunque las etiquetas de los extremos sean idénticas en ambos formatos. Este efecto se ve reducido cuando la escala incluye un punto intermedio neutral, que tiende a acumular una mayor frecuencia de respuestas. Esta es, de hecho, una característica general, ya que, aunque las personas encuestadas tienden a distribuir sus respuestas entre todas las categorías, se suele observar una cierta regresión a la media. Aunque también es frecuente encontrar el efecto de los puntos de referencia, según el cual ciertos números o etiquetas se asocian con ciertos significados que pueden influir en la probabilidad de que las personas encuestadas seleccionen, o no, una determinada respuesta.

Según Van Vaerenberh y Thomas (2012), hay varios estilos de respuesta. Los que más atención han recibido son, a grandes rasgos, los mismos identificados por Toruangeau et al. (2000). En concreto, se trata del estilo de respuesta de aquiescencia (ERA), que es la tendencia a estar de acuerdo con los ítems independientemente del contenido; el estilo de respuesta de punto medio (ERM), que es la tendencia a usar la categoría de respuesta media de una escala de calificación independientemente de los contenidos; el estilo de respuesta extremo (ERE) o tendencia a utilizar las categorías de respuesta más altas y más bajas de una escala de calificación; y estilo de respuesta suave (ERS), que refleja la tendencia a evitar las categorías de respuesta más altas y más bajas, complementando así al ERE. Por otro lado, Toepoel et al. (2008) han encontrado que la categoría intermedia (cuando está presente) es la respuesta seleccionada con mayor frecuencia; además, las respuestas en un formato de respuesta de negativo a positivo difieren significativamente de las obtenidas cuando el formato va de positivo a negativo.

Los estilos de respuesta deben tenerse en cuenta ya que la opinión manifestada al responder a una encuesta refleja tanto la opinión verdadera como el efecto de errores aleatorios y sistemáticos. Los estilos de respuesta son la tendencia de los encuestados a responder a las preguntas de la encuesta de cierta manera, independientemente del contenido, y contribuyen a un error sistemático que afecta tanto a las distribuciones univariadas (distribuciones de frecuencia, medias y variaciones) como a las distribuciones multivariadas, ya que afecta la magnitud de las correlaciones entre las variables. Por tanto, muchas técnicas estadísticas, como el alfa de Cronbach, el análisis de regresión, el análisis factorial y el modelado de ecuaciones estructurales se basan en correlaciones entre variables. Como resultado, los estudios que examinan tales relaciones sin controlar los estilos de respuesta pueden arrojar resultados engañosos (Van Vaerenberh y Tomas, 2012).

De acuerdo con Peabody (1962) o Komorita (1963), al medir actitudes pueden distinguirse dos componentes: 1) la dirección de la actitud (positiva o negativa) y 2) la intensidad de la actitud. No obstante, no resulta del todo evidente si diferenciar entre ambos componentes es necesario o deseable. De acuerdo con estos autores, cuando los distintos ítems se suman, los resultados obtenidos reflejan de manera fundamental la dirección de las respuestas, y en muy poca medida, la intensidad de la actitud. La implicación práctica está clara para estos autores: hay que utilizar ítems dicotómicos (dos opciones de respuesta) que reflejen la dirección de la actitud, pues las diferencias en la intensidad parecen ser de manera mayoritaria el resultado de los diferentes estilos de respuesta. Por otro lado, hay evidencia de que existe una relación curvilínea entre la fiabilidad test-retest en las encuestas que miden actitudes y el número de opciones. Las escalas con dos categorías ofrecen mejores resultados que las que tienen tres. Con cuatro o más opciones, la fiabilidad aumenta, aunque a un ritmo decreciente (Alwin, 1992).

### **1.7 Algunas consideraciones metodológicas**

Hay varias cuestiones interesantes que se pueden destacar a raíz de la revisión realizada de la literatura publicada al respecto que, sin ser exhaustiva, cubre un amplio periodo de tiempo. En primer lugar, es de destacar que a pesar de ser un tema analizado desde los inicios de la investigación con encuestas, en torno a los años 40-50 del siglo pasado, sigue sin estar claro cuál es la escala que proporciona mejores resultados en la medición de actitudes y opiniones.

En segundo lugar, la investigación sobre la identificación del número óptimo de alternativas de respuesta en escalas A/D se ha centrado principalmente en analizar la fiabilidad. Siempre que se medie algo, la medición obtenida incluye una cierta cantidad de error aleatorio. El objetivo central del análisis de la fiabilidad es estimar el error cometido. En el modelo lineal clásico, que es el que se ha aplicado en los estudios con encuestas, se parte de la premisa de que la puntuación empírica de un sujeto en una prueba consta de dos componentes, la puntuación verdadera, la que le corresponde en esa escala, dadas sus características, y un componente de error aleatorio. El objetivo de la fiabilidad consiste en estimar el error, de tal manera que a partir de él se pueda identificar la puntuación verdadera (Muñiz, 1996).

La fiabilidad, por tanto, se aplica a escalas más que a ítems, y se suele definir como la capacidad de la escala para medir de forma precisa y consistente la característica que se pretende medir. Hay distintas formas de analizar la fiabilidad. En los estudios con encuestas, los métodos empleados con más frecuencia han sido la fiabilidad test-

retest, que implica correlacionar las puntuaciones obtenidas en dos ocasiones por los mismos sujetos en la misma prueba (o, en el caso que nos ocupa, en las mismas preguntas) y hace referencia a la estabilidad de las puntuaciones. También se ha utilizado, en menor medida, el análisis de la consistencia interna, que depende del grado en que los elementos del test covarían entre sí y, por tanto, refleja la capacidad de una escala (no se puede aplicar a ítems individuales por su propia concepción) para medir lo que se quiere medir de manera homogénea (Muñiz, 1996).

No obstante, la realidad es que la medición de opiniones y actitudes en las encuestas de opinión no se ajusta bien a la medición de la consistencia interna, pues estas encuestas suelen incluir un número reducido de preguntas (o ítems) para medir la opinión sobre cada tema de interés. Tampoco es habitual que se pase de forma repetida, en un periodo breve de tiempo, la misma pregunta a los mismos sujetos, aunque esta es una realidad que está cambiando con la tendencia cada vez mayor a constituir paneles, grupos de personas que, voluntariamente o a cambio de una pequeña gratificación dineraria o en especie, se ponen a disposición de las empresas de estudios de mercado y opinión para responder a las encuestas y/o preguntas que estas requieren. En todo caso, la evidencia analizada indica que la consistencia interna es independiente del número de alternativas empleadas (Mattel y Jacoby, 1972). Si atendemos a la estabilidad, los resultados obtenidos hasta la fecha, como se ha señalado, apuntan a una relación curvilínea, de tal manera que la estabilidad de las puntuaciones aumenta al hacerlo el número de opciones, aunque el resultado es peor para las escalas de tres puntos que para las de dos.

Por otro lado, igual que la calidad de un instrumento para medir una magnitud física depende de la capacidad para realizar mediciones precisas y, por tanto, para discriminar entre pequeñas variaciones, sería deseable que las encuestas de opinión puedan medir las actitudes de la población de manera precisa y, por tanto, sería interesante encontrar cuál es el formato de la escala de respuesta que mejor contribuye a discriminar entre distintas opiniones y actitudes. Además, aunque hay evidencia desde los años 40 del siglo pasado que muestra el efecto del formato de respuesta sobre las respuestas obtenidas en encuestas de opinión, no se ha analizado si este efecto genera resultados cuantitativamente diferentes; es decir, ¿depende la intensidad de las actitudes de las categorías de la escala de respuesta con las que se miden? En este sentido, con el objetivo de analizar la posible influencia del formato de la escala de respuesta en las medias y desviaciones típicas de las puntuaciones, Leung (2011) transformó las puntuaciones en una escala de 0 a 100. Una vez realizada la transformación, no encontró que hubiera diferencias significativas en las

medias ni en las desviaciones típicas entre escalas de 4, 5, 6 y 11 categorías que, para garantizar la comparabilidad, solo se etiquetaron en los extremos. También analizó si las respuestas se distribuían normalmente. El resultado fue afirmativo para las escalas de 6 y 11 puntos, pero no para las de 4 y 5. Tampoco encontró diferencias en función de que la escala tuviera, o no, un punto intermedio neutral.

## **1.8 Objetivo**

Teniendo en cuenta todo lo señalado hasta el momento, este trabajo plantea dos objetivos. Por un lado, analizar si ciertas características del formato de la escala de respuesta contribuyen a obtener estimaciones diferentes de la intensidad de las actitudes medidas. Por otro, si estas características influyen en la capacidad de discriminación de las preguntas diseñadas para medir actitudes.

## **2. DESARROLLO DEL TRABAJO**

### **2.1 Diseño del experimento**

Las encuestas de opinión suelen medir las opiniones/actitudes mediante una escala bipolar de A/D con cinco categorías y un punto intermedio neutral (ni de acuerdo, ni en desacuerdo) a pesar de que suscita importantes dudas metodológicas como ya se ha señalado (p. e. Revilla et al. (2013)). Teniendo esto en cuenta, se diseñó un experimento para poner a prueba distintos formatos de respuesta que permitiera comparar la escala de cinco opciones con las escalas pares adyacentes, es decir, se incluyeron distintos formatos de escala de respuesta con 4, 5 y 6 categorías. En concreto, el experimento incluyó 8 formatos distintos que variaron en función de 6 seis características que han permitido definir seis variables independientes: 1) “opciones” con tres valores, 4, 5 y 6, en función del número de categorías de la escala de respuesta; 2) “parimpar” que agrupa las opciones 4 y 6 por un lado (número par de categorías) y asigna otro valor a la opción con 5 categorías (impar); 3) “etiquetas” con dos valores según todas las categorías de respuesta tengan etiquetas verbales para definir las o solo tengan etiquetas en los extremos; 4) “polaridad” para definir si la escala es unipolar (solo acuerdo o solo desacuerdo) o bipolar (mide tanto acuerdo como desacuerdo); 5) “extremo” para reflejar si el extremo de la escala incluye el término “muy” o “totalmente”; y 6) “orden” para especificar si la escala está organizada desde de acuerdo hacia desacuerdo, o viceversa.

El experimento incluyó 8 ítems agrupados en dos preguntas. Los cuatro elementos o ítems de la pregunta 1 hacen referencia directamente a la persona que responde (“las asignaturas de ciencia siempre se me dieron mal”, “la ciencia es tan especializada que

me cuesta entenderla”, “considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana” y “la información que oigo acerca de la ciencia suele ser cierta”). Los ítems de la pregunta 2 hacen referencia a cuestiones generales sobre la ciencia, o sobre los ciudadanos en general, pero no están focalizados en el sujeto que responde (“la investigación científica depende cada vez más de intereses comerciales”, “el desarrollo científico y tecnológico puede tener consecuencias indeseadas”, “los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos” y “es poco probable que los ciudadanos llegemos a saber realmente cuáles son los riesgos de los avances científicos”). Esta decisión se basó en dos consideraciones. Por un lado, es bien conocido que las personas encuestadas tienen mayor propensión a generar estilos de respuesta particulares cuando se enfrentan a conjuntos amplios de ítems que utilizan la misma escala de respuesta (p. e. Krosnick, 1991); por tanto, se consideró oportuno no juntar los ocho ítems en la misma pregunta. Por otro, en la investigación con encuestas se pone el foco en los encuestados y, por tanto, en las características o rasgos que los hacen responder a una encuesta de una forma u otra (p. e. Krosnick, 1991; Schuman y Presser, 1996). Pero no hay duda de que la dificultad de las preguntas interacciona con las características de los sujetos. Desde esta perspectiva, Shoemaker et al. (2002) pidieron a una muestra de estudiantes graduados que evaluaran la dificultad de los ítems actuando como “expertos”. Los autores encontraron que cuanto más esfuerzo cognitivo era necesario involucrar en responder a una pregunta, mayor era el número de respuestas no significativas (no sabe /no contesta). Por tanto, detrás de la elección de los dos conjuntos de ítems divididos en las preguntas 1 y 2 está el deseo de analizar si el efecto del formato de la escala de respuesta se ve influido por la dificultad de las preguntas.

Una vez definidos los formatos de respuesta y los ítems, la Unidad de Desarrollo de Aplicaciones y Sistemas Informáticos del CIEMAT desarrolló una aplicación para administrar la encuesta por Internet de tal manera que, cada vez que alguien hacía clicaba en el enlace para responder a la encuesta, la aplicación seleccionaba aleatoriamente uno de los 8 formatos de respuesta para cada una de las preguntas. La decisión de no seleccionar el mismo formato para las dos preguntas, no obstante, ha resultado errónea, pues no permite comparar de manera directa las respuestas de las personas participantes a las dos preguntas (no son muestras relacionadas).

Se crearon dos enlaces distintos. Uno se distribuyó a la red de colaboradores que ha generado la Unidad de Investigación en Cultura Científica en los distintos estudios y experimentos con encuestas que ha realizado en los últimos años; también se

distribuyó en las redes sociales de esta unidad para captar voluntarios que participaran en el estudio. El segundo enlace se distribuyó entre los trabajadores y trabajadoras del CIEMAT. Como resultado, se ha creado una séptima variable independiente, denominada muestra, que diferencia entre ambos grupos de participantes.

Por último, como se ha señalado repetidamente, las encuestas sitúan a quien responde en la tesitura de formular juicios y formarse opiniones que debe manifestar a través de las opciones de respuesta que se le proporcionan. Sin embargo, al analizar las respuestas a estas encuestas se ha tendido a ignorar las diferencias individuales. Y las diferencias en las habilidades o rasgos cognitivos desempeñan un papel fundamental en los juicios y opiniones. Para poder analizar la influencia de estas diferencias, se ha incluido en el cuestionario un ítem del Test de Reflexión Cognitiva (TRC). El TRC fue introducido en la literatura por Frederick (2005) y está formado por tres ítems diseñados para medir la tendencia a ignorar la respuesta automática, incorrecta, e implicarse en un análisis más profundo que conduce a la respuesta correcta. Es decir, mide la capacidad del sistema 2 para imponerse al sistema 1. La respuesta a este problema se ha incluido en el fichero de datos como otra variable independiente con el nombre de “problema” y tres opciones: solución incorrecta, solución correcta y abandono (falta de respuesta).

El cuadernillo con los ítems, los formatos de respuesta y la codificación de todas las variables incluidas en el estudio está disponible en el Anexo I.

## **2.2 Ajuste del fichero de datos**

Para realizar este estudio hemos partido de un fichero SPSS que contenía 66 variables: 8 opciones de formato x 4 ítems x 2 preguntas, más las variables muestra y problema. Cada una de las 64 variables que recogían las respuestas de los sujetos incluía solo los datos de quienes habían respondido a esa opción, el resto eran NA. Por tanto, una cuestión fundamental para realizar el análisis estadístico ha sido el ajuste del fichero de datos. En primer lugar se ha realizado una depuración de los nombres de las variables para poder realizar posteriormente el tratamiento de los datos con objeto de obtener las variables de interés (variables independientes y dependientes). A continuación se pasó a formato largo con la función “reshape” para tener todas las opciones del formato de respuesta y los distintos ítems de cada pregunta en una variable y, de este modo, eliminar los NA. El siguiente paso fue el diseño de la función “tratar”, que ha permitido generar un fichero de datos con las 6 variables independientes relacionadas con el formato de respuesta y las variables que se necesitan para definir adecuadamente las variables dependientes. El siguiente paso

consistió en la creación de las dos variables dependientes. En primer lugar, se ha definido una variable denominada “variabilidad” que contabiliza el número de respuestas distintas a los cuatro ítems de cada pregunta y lo divide entre el número de ítems. En segundo lugar se ha definido otra variable, denominada “transformadas” que supone la transformación de las puntuaciones brutas en una escala de 0 a 10 para, de este modo, poder comparar las medias de las respuestas de los sujetos en función de las variables independientes que recogen las características de los distintos formatos de respuesta. Por último, el fichero largo se pasó a formato ancho. Este es el fichero con el que se ha realizado el análisis de datos. Incluye las seis variables independientes creadas con la función “tratar, las dos variables independientes que no han necesitado ninguna intervención, una variable dependiente que es la suma de las puntuaciones transformadas en una escala de 0 a 10 a los cuatro ítems de cada pregunta (“sumT”) y otra variable dependiente que representa el índice de variación (“variabilidad”) para cada sujeto en cada pregunta. El código desarrollado en este proceso figura en el Anexo II.

### **2.3 Análisis estadístico**

Como se ha señalado en el apartado 2.1, al diseñar la aplicación se tomó la decisión de que la selección aleatoria de la variedad del formato de respuesta fuera independiente para la pregunta 1 y para la pregunta 2; la consecuencia directa es que la mayor parte de los sujetos respondieron a ambas preguntas utilizando formatos distintos. Por este motivo, el análisis de la posible influencia de las variables independientes sobre las dos variables dependientes debe hacerse por separado para cada pregunta y las comparaciones entre ambas se tienen que producir a posteriori a partir de los resultados obtenidos en los análisis por separado. Por tanto, para analizar el efecto de las ocho variables independientes sobre las dos variables dependientes realizaremos ocho análisis de diferencia de medias con muestras independientes para cada pregunta.

En primer lugar se realiza el análisis descriptivo: obtención del resumen estadístico de la variable dependiente en función de los grupos de las variables independientes con la función “describeBy” del paquete “psych” y representación gráfica mediante el paquete “ggplot2”. El intervalo en el que están comprendidos los valores de la variable “sumT” es 0-40 y, por tanto, se puede considerar que es una variable cuantitativa discreta con un número suficiente de valores distintos para realizar la representación gráfica mediante histograma para la variable sola y utilizando el gráfico de cajas para representar los valores en función de las variables independientes. Por el contrario, al

disponer solo de cuatro ítems por pregunta en el diseño del experimento, la variable “variabilidad” es discreta con cuatro valores y se necesita representar gráficamente mediante el gráfico de barras.

En segundo lugar, con el fin de elegir la mejor prueba estadística, se han puesto a prueba los supuestos en los que se basan las técnicas paramétricas que permiten analizar la diferencia de medias: que no haya outliers, que los distintos grupos se distribuyan normalmente (supuesto de normalidad) y que tengan la misma varianza (supuesto de homocedasticidad). La identificación de valores extremos se ha realizado con la función “rp.outlier” del paquete “rapportools”. Para comprobar el supuesto de normalidad se ha utilizado la función “shapiro.test”. Como se ha rechazado la hipótesis de normalidad en todos los análisis realizados, para comprobar el supuesto de homocedasticidad se ha utilizado el test de Fligner-Killeen (función “fligner.test”).

En función de los resultados obtenidos, para realizar el análisis sobre la diferencia de medias se han utilizado algunas de las pruebas robustas implementadas en la librería “WRS2” (funciones “yuen” y “t1way”, con la función “lincon” para realizar las comparaciones múltiples) y las pruebas no paramétricas de Wilcoxon y de Kruskal-Wallis (funciones “wilcox.test” y “kruskal.test” con la función “kruskalmc” de la librería “pgirmess” para realizar las comparaciones múltiples).

El código con los análisis realizados está incluido en los Anexos III a VII.

## **2.4 Hipótesis**

H0: La media de las puntuaciones transformadas es igual para los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)

H1: La media de las puntuaciones transformadas difiere entre los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema).

## **2.5 Procedimiento**

Se ha dividido el fichero definitivo (“excope\_18”, experimento sobre cómo preguntar, edición de 2018) en dos, separando a partir de la pregunta y se ha realizado el análisis para los datos de cada una de ellas por separado. Se ha separado también el resultado en función de la variable dependiente analizada. En total, se han realizado cuatro bloques de análisis, uno por pregunta y variable dependiente. En cada uno de ellos, el procedimiento ha sido el siguiente: realización del análisis descriptivo,

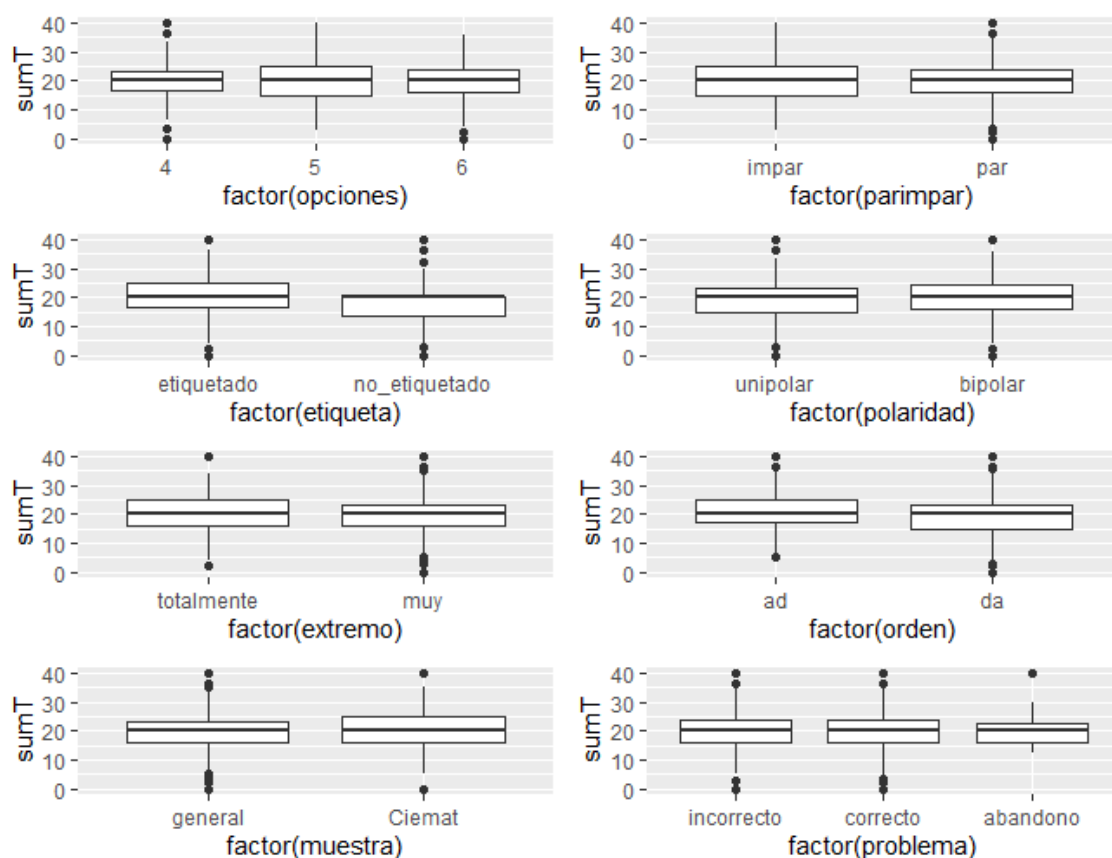
comprobación de los supuestos (identificación de outliers, comprobación del supuesto de normalidad y comprobación del supuesto de homocedasticidad), análisis de la diferencia de medias y, en el caso de encontrar diferencias estadísticamente significativas, análisis de la influencia conjunta de las variables que han proporcionado un resultado estadísticamente significativo. El código empleado se recoge en los Anexos III a VI.

## 2.6 Resultados

### A) Variable dependiente: suma de las puntuaciones transformadas (sumT)

Los principales resultados del análisis de la variable dependiente suma de las puntuaciones transformadas en una escala de 0 a 10 (sumT) se recogen en la tabla 1. Los gráficos de cajas están incluidos en las figuras 1a y 1b.

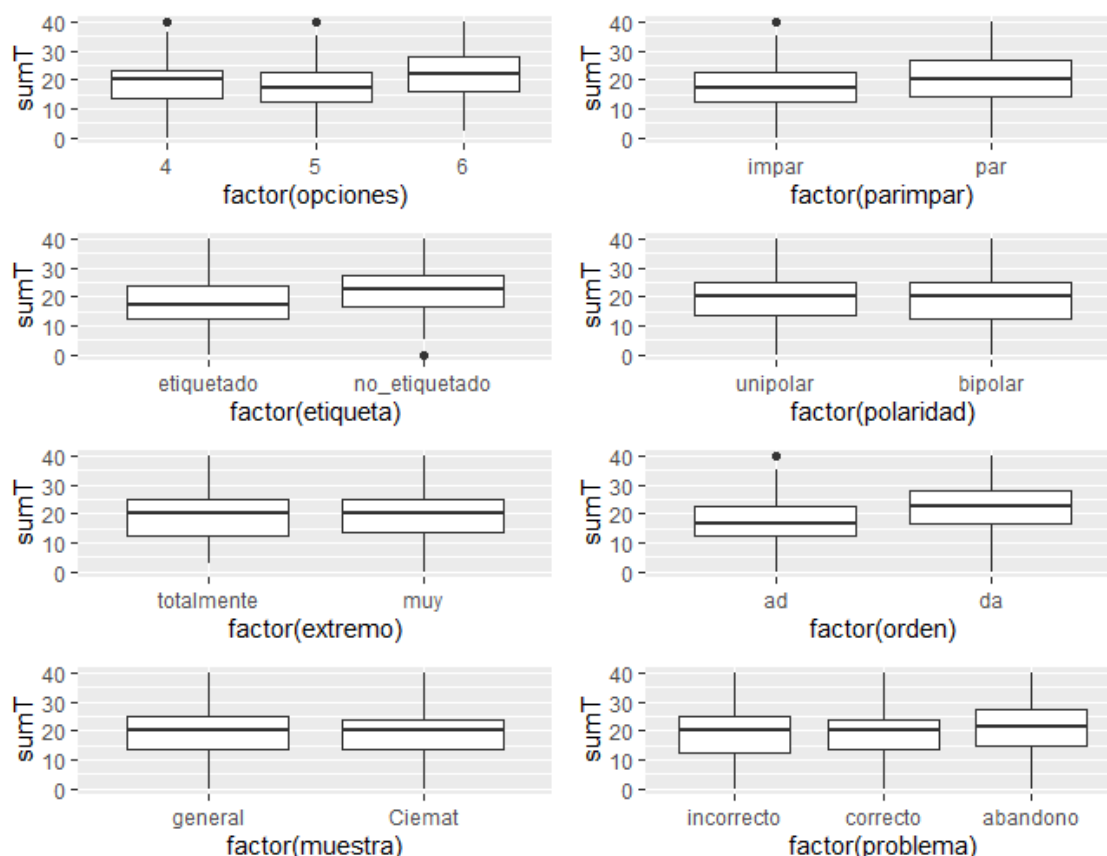
Figura 1a. Gráficos de caja – Variable “sumT”. Pregunta 1



En todos los gráficos de la P1 y la mayoría de la P2 hemos encontrado valores extremos, aunque están muy próximos a los bigotes, porque, como se puede ver en el histograma de la figura 3, la variable se distribuye en forma de campana entre los valores mínimo y máximo sin valores extremos, que tampoco hemos identificado en la mayor parte de los análisis realizados con la función “rp.outlier”. Las excepciones han

llevado a la identificación de dos sujetos que han elegido la opción más extrema en los cuatro ítems de la pregunta 1 en los grupos “no etiquetado” de la variable “orden” y “abandono” de la variable “problema”. En todos los análisis realizados hemos rechazado el supuesto de normalidad, pero la evidencia ha llevado a mantener el supuesto de heterocedasticidad, excepto en los grupos formados a partir de la variable “problema” en la pregunta 2. Por lo tanto, en el análisis de la relación entre “sumT” y las variables independientes se han utilizado mayoritariamente las pruebas robustas.

Figura 1b. Gráficos de caja – Variable “sumT”. Pregunta 2



En la pregunta 1 solo se han encontrado diferencias estadísticamente significativas en la variable sumT debidas al etiquetado o no de las categorías de respuesta y al orden en el que se han presentado las opciones de respuesta (ad, de acuerdo a en desacuerdo; da, en desacuerdo a de acuerdo). Los resultados en la pregunta 2 han sido algo diferentes. Se han encontrado diferencias estadísticamente significativas debidas al número de opciones, de tal manera que se han obtenido puntuaciones más altas en las escalas con 6 opciones de respuesta en comparación con las de 4 y 5. También se han encontrado diferencias en función de la polaridad de la escala, siendo mayores cuando las escalas de respuesta son pares (4 y 6 opciones de respuesta, en comparación con 5 opciones de respuesta). Por último, tal y como ha ocurrido con la

pregunta 1, en la pregunta 2 se han encontrado diferencias relacionadas con el etiquetado de las opciones y el orden de presentación, pero en sentidos contrarios. Es decir, en la pregunta 1 se han obtenido puntuaciones significativamente más altas cuando las categorías estaban etiquetadas y cuando el orden era de acuerdo hacia desacuerdo. En la pregunta 2 ha ocurrido justo lo contrario. La puntuación ha tendido a ser más alta cuando no había etiquetas y el orden era desde desacuerdo hacia acuerdo.

Tabla 1. Análisis de la influencia de las variables independientes en la variable “sumT”

VD: sumT				
VI	P1		P2	
Opciones	Media(SD) - n	p	Media(SD) - n	p*
4	19,69 (7,02) - 194	> 0,05	18,99 (8,33) - 217	< 0,01 T.E.: 0,29
5	20,03 (5,79) - 457		17,91 (7,98) - 455	
6	20,19 (6,02) - 217		21,99 (7,92) - 201	
* En P2, la media es significativamente mayor en el grupo 6 en comparación con los grupos 4 y 5				
Parimpar	Media(SD) - n	p	Media(SD) - n	p
par	20,03 (5,79) - 457	> 0,05	17,91 (7,98) - 418	< 0,01 T.E.: 0,24
impar	19,96 (6,51) - 411		20,43 (8,26) - 418	
Etiqueta	Media(SD) - n	p	Media(SD) - n	p
etiquetado	20,55 (6,09) - 663	< 0,01	18,11 (8) - 659	< 0,01
no etiquetado	18,19 (5,93) - 205	T.E.: 0,32	22,22 (8,06) - 214	T.E.: 0,34
Polaridad	Media(SD) - n	p	Media(SD) - n	p
unipolar	19,66 (6,19) - 400	> 0,05	19,26 (8,26) - 436	> 0,05
bipolar	20,28 (6,08) - 468		18,97 (8,16) - 437	
Extremo	Media(SD) - n	p	Media(SD) - n	p
totalmente	20,17 (6,32) - 234	> 0,05	18,62 (8,09) - 205	> 0,05
muy	19,93 (6,07) - 634		19,27 (8,24) - 668	
Orden	Media(SD) - n	p	Media(SD) - n	p
a-d	20,72 (6,13) - 446	< 0,01	16,4 (7,43) - 458	< 0,01
d-a	19,22 (6,05) - 422	T.E.: 0,21	22,11 (7,98) - 415	T.E.: 0,49
Muestra	Media(SD) - n	p	Media(SD) - n	p
general	19,74 (6) - 554	> 0,05	19,12 (8,66) - 555	> 0,05
Ciemat	120,44 (6,36) - 314		19,12 (7,36 - 318)	
Problema	Media(SD) - n	p	Media(SD) - n	p
Incorrecto	20,24 (6,4) - 367	> 0,05	19,05 (8,72) - 368	> 0,05
Correcto	19,8 (5,97) - 462		19,04 (7,67) - 465	
Abandono	19,98 (5,64) - 39		20,53 (9,29) - 40	
Media(SD) - n	19,99(6,14) - 868		19,12 (8,21) - 873	

Como se acaba de señalar, se han encontrado diferencias significativas en el análisis de la variable “sumT” debidas al orden y a la etiqueta. Hay que tener en cuenta, no obstante, que hay una relación entre estas dos variables independientes. En concreto, en la pregunta 1, de los 446 sujetos que forman parte del grupo “ad”, en la variable “orden”, es decir, los que han respondido a una escala ordenada desde “acuerdo”

hacia “desacuerdo”, ninguno de ellos ha respondido a una escala no etiquetada. En cambio, 217 de los que han respondido en el orden inverso lo han hecho a una escala etiquetada, y 205 a una escala no etiquetada. Por lo que respecta a la pregunta 2, ninguno de los 458 sujetos que han respondido en el orden “ad” ha recibido una escala no etiquetada; sin embargo, 214 de los que contestaron en una escala en orden “da” lo hicieron en una escala no etiquetada y 201 en una etiquetada. Por tanto, para analizar si el efecto encontrado en la relación entre la VD y el orden en realidad se debe a la etiqueta, hemos seleccionado a los sujetos que han respondido a la escala en el orden “da” y hemos analizado si hay diferencias debidas a la etiqueta de la escala de respuesta. El resultado indica que hay diferencias debidas a la presencia o no de etiqueta en las categorías de respuesta. En este caso volvemos a encontrar resultados diferentes en las dos preguntas. En la P1, hay diferencias en las respuestas debidas tanto al orden como a la presencia de etiquetas. En la P2, la diferencia encontrada en la variable “etiqueta” es, en realidad, debida a la variable “orden”; por tanto, cuando analizamos si hay diferencias debidas a la etiqueta entre los sujetos que han respondido en orden “da” no obtenemos evidencia para rechazar la hipótesis nula. El tamaño del efecto encontrado en la pregunta 1 es medio bajo (igual a 0,24).

### **B) Variable dependiente: variabilidad**

Como ya se ha señalado, “variabilidad” es una variable discreta con cuatro valores, debido al hecho de que solo tenemos 4 ítems. Por tanto, el gráfico de cajas no es apropiado para representar esta variable y utilizaremos el diagrama de barras. Además, por ser una variable discreta, utilizaremos las pruebas no paramétricas, a no ser que se incumpla el supuesto de homocedasticidad, a cuyo incumplimiento son muy sensibles estas pruebas. Si este caso se diera, utilizaremos las pruebas robustas. Los principales resultados se muestran en la tabla 2. Los diagramas de barras, en las figuras 2a y 2b.

En ambas preguntas hay diferencias significativas en la variable “variabilidad” debidas al número de opciones. En P1, la variabilidad es significativamente menor en el grupo de 4 opciones en comparación con los otros dos. En P2, hay mayor variabilidad en el grupo de 6 opciones en comparación con los otros dos. El tamaño del efecto es igual a 0,23 y 0,24 respectivamente. A la hora de valorar la magnitud del efecto, conviene tener en cuenta, como se ha señalado, que “variabilidad” tiene muy poca variación debido al reducido número de ítems; por tanto, se puede considerar que un tamaño del efecto de 0,23 es relevante. Al igual que ocurre con la variable “sumT”, el orden en el que se presentan las opciones de respuesta influye en la variabilidad de las

respuestas ofrecidas por los participantes. En este caso, a diferencia de lo que ha ocurrido con la variable “sumT”, la variabilidad es mayor en el orden “da” en ambas preguntas. Se observan también diferencias relacionadas con la respuesta al problema del TRC en la pregunta 2; en concreto, las respuestas de los que no han proporcionado una solución al problema son significativamente menos variadas.

Tabla 2. Análisis de la influencia de las variables independientes en la variable “variabilidad”.

VD: sumT				
VI	P1		P2	
Opciones	Media(SD) - n	p*	Media(SD) - n	p**
4	0,7 (0,19) - 194	< 0,01 T.E.: 0,24	0,59 (0,18) - 217	< 0,01 T.E.: 0,23
5	0,74 (0,16) - 457		0,61 (0,19) - 455	
6	0,75 (0,17) - 217		0,65 (0,19) - 201	
* En P1, la variabilidad es significativamente menor para 4 opciones respecto a 5 y 6; ** en P2, menor variabilidad en el grupo de 4 opciones				
Parimpar	Media(SD) - n	p	Media(SD) - n	p
par	0,74 (0,16) - 457	> 0,05	0,61 (0,19) - 455	> 0,05
impar	0,73 (0,18) - 411		0,62 (0,19) - 418	
Etiqueta	Media(SD) - n	p	Media(SD) - n	p
etiquetado	0,73 (0,17) - 663	> 0,05	0,61 (0,19) - 659	> 0,05
no etiquetado	0,74 (0,17) - 205		0,61 (0,19) - 214	
Polaridad	Media(SD) - n	p	Media(SD) - n	p
unipolar	0,72 (0,17) - 400	> 0,05	0,61 (0,18) - 436	> 0,05
bipolar	0,74 (0,17) - 468		0,62 (0,19) - 437	
Extremo	Media(SD) - n	p	Media(SD) - n	p
totalmente	0,74 (0,17) - 234	> 0,05	0,61 (0,18) - 205	> 0,05
muy	0,73 (0,17) - 634		0,61 (0,19) - 668	
Orden	Media(SD) - n	p	Media(SD) - n	p
a-d	0,72 (0,17) - 446	< 0,01	0,59 (0,18) - 458	< 0,01
d-a	0,75 (0,17) - 422		0,63 (0,19) - 415	
Muestra	Media(SD) - n	p	Media(SD) - n	p
general	0,74 (0,17) - 554	> 0,05	0,61 (0,19) - 555	> 0,05
Ciemat	0,72 (0,18) - 314		0,62 (0,19) - 318	
Problema	Media(SD) - n	p	Media(SD) - n	p***
Incorrecto	0,73 (0,18) - 367	> 0,05	0,62 (0,18) - 368	< 0,01
Correcto	0,74 (0,17) - 462		0,61 (0,19) - 465	
Abandono	0,7 (0,19) - 39		0,52 (0,19) - 40	
*** Significativamente menor variabilidad entre los que no han proporcionado una solución				
<b>Media(SD) - n</b>	<b>0,73 (0,17) - 868</b>		<b>0,61(0,19) - 873</b>	

Figura 2a. Gráficos de barras – Variable “variabilidad”. Pregunta 1

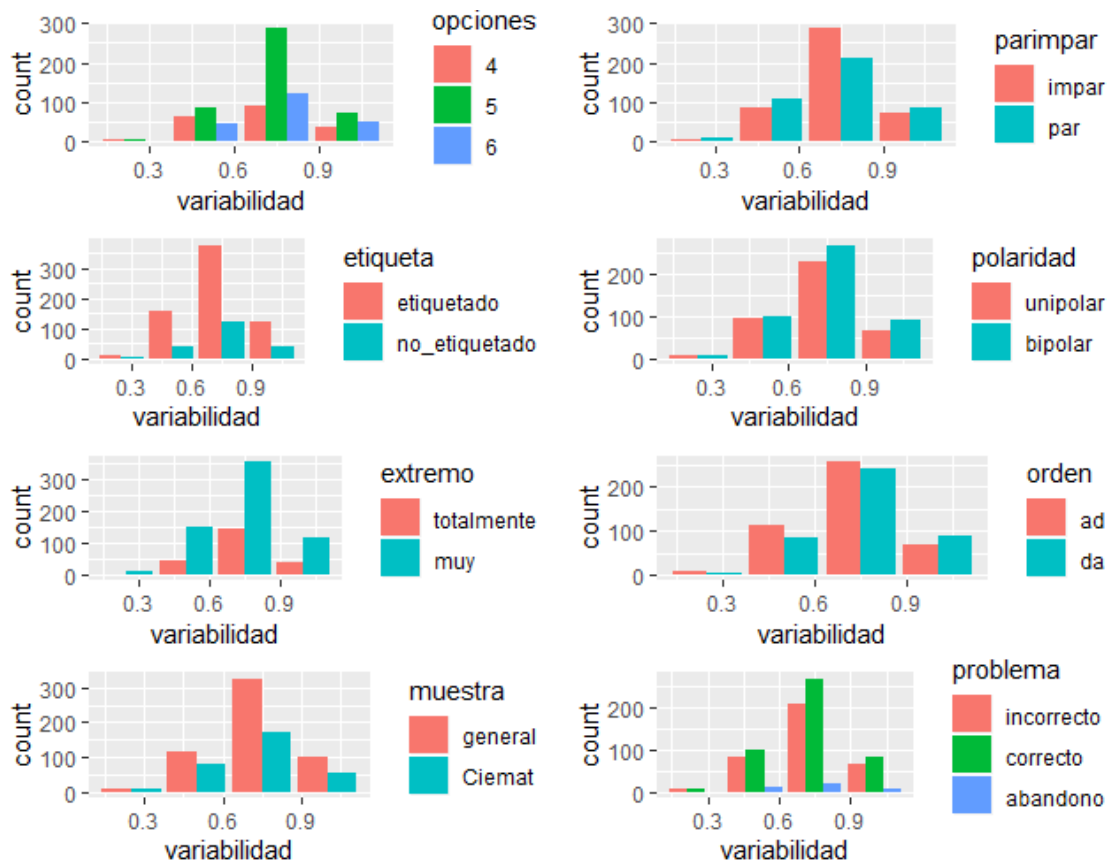
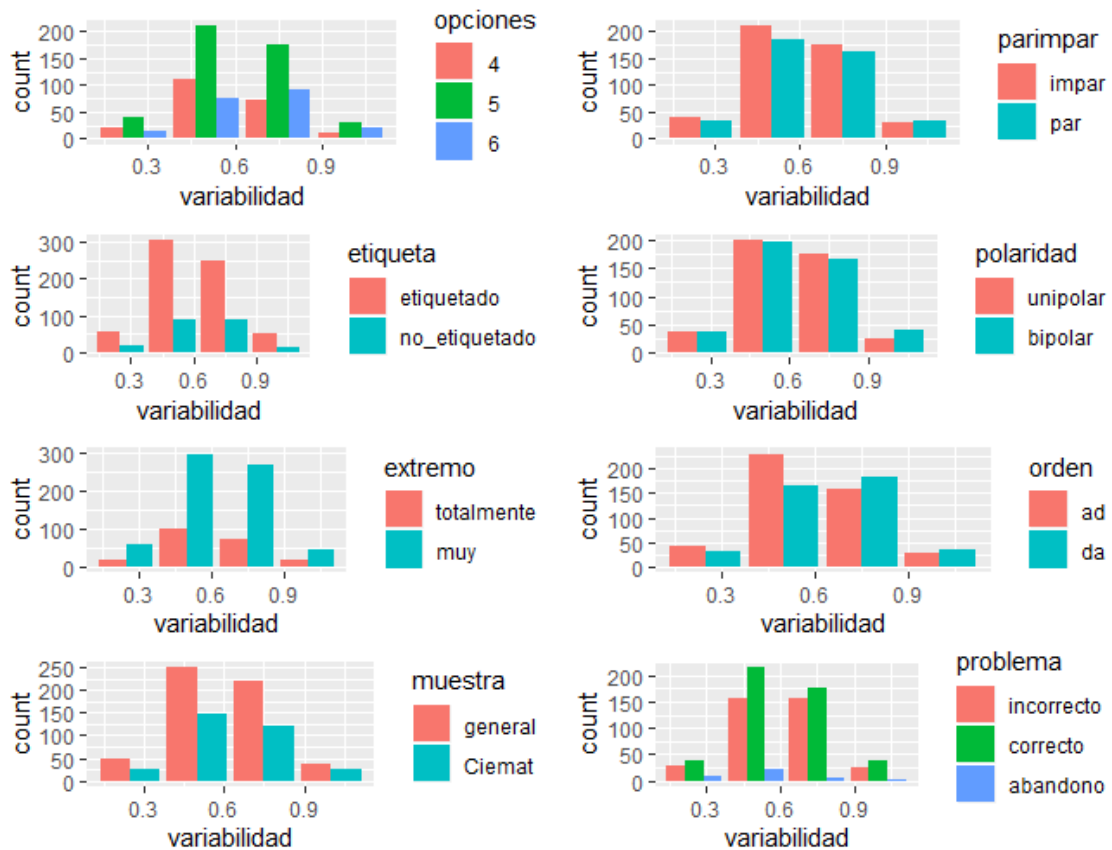


Figura 2b. Gráficos de barras – Variable “variabilidad”. Pregunta 2



En ninguno de los análisis realizados hemos encontrado evidencia de que haya diferencias en la forma de reaccionar ante los distintos formatos de respuesta debidas a la procedencia de la muestra, la población general y personas pertenecientes al Ciemat, que es un Organismo Público de Investigación y, por tanto, en el que la ciencia ocupa un lugar central. No obstante, es importante tener en cuenta que la muestra de la población general en realidad está formada por voluntarios que forman parte de la red de colaboradores de la UICC del Ciemat por un lado, o a los que se ha contactado a través del Twitter de la propia unidad o de sus integrantes.

El efecto del orden de presentación de las opciones es interesante. En la variable “sumT” y en la pregunta 2, se observa que la puntuación media es mayor cuando el orden va desde desacuerdo hacia acuerdo. Este resultado en realidad estaría indicando que no hay efecto, puesto que al ir en orden inverso, una puntuación más alta en el orden “da” indica un mayor acuerdo, lo mismo que indica una puntuación más baja en el orden “ad”. Es el mismo efecto que muestra la variable “etiqueta”:\_ la presencia de etiquetado genera puntuaciones más altas en la P1, mientras que genera puntuaciones más bajas en la P2. El efecto de la etiqueta se mantiene a pesar del orden en la P1, como ya se ha señalado, pero no lo hace en la P2. Estos resultados indican una interacción entre el efecto del formato de respuesta y el contenido de la pregunta.

### **C) Análisis de las diferencias entre las preguntas 1 y 2**

Como se ha señalado en el epígrafe en el que se describe el diseño del experimento, al separar entre las preguntas 1 y 2 se asumió que la pregunta 2 era más difícil de contestar pues hacía referencia a cuestiones “externas” a la persona y centradas en aspectos concretos de la ciencia sobre los que podía no tener una opinión formada. Para poner a prueba esta hipótesis, se contactó con 99 de los participantes en el experimento para pedirles, como tarea adicional, que evaluaran la dificultad de la tarea que se les había pedido. En concreto, se les proporcionó la siguiente información:

*Uno de los problemas de las encuestas de opinión es que dan por supuesto que las personas tenemos una opinión formada sobre cualquier cosa sobre la que los investigadores nos quieran preguntar. Pero no siempre es así y, por tanto, hay preguntas que nos parecen más difíciles de contestar que otras.*

*Ejemplos:*

*A. Los efectos indeseados del fracking superan con creces a los beneficios*

*B. La ciencia es difícil*

*Parece claro que el enunciado B es más fácil, porque solo tenemos que mirarnos a nosotros mismos para poder opinar, mientras que en el enunciado A necesitamos saber qué es el fracking, cuáles son los efectos indeseados, cuáles los beneficios y ponderarlos.*

*Teniendo esto en cuenta, necesitamos que valore la dificultad de 10 preguntas, utilizadas ya en una encuesta, utilizando una escala de cinco puntos: mucho, bastante, algo, poco o muy poco.*

Al analizar la dificultad atribuida a los ocho primeros ítems, que son los que forman parte de este experimento, se encontró que había diferencias significativas en la dificultad atribuida a los dos grupos de ítems (1 a 4 y 5 a 8) (dificultad de P1 – media igual a 2,03 y desviación típica igual a 0,73; dificultad de P2 – media igual a 2,74 y desviación típica igual a 0,93;  $t(98) = -7,33$ ,  $p < 0,01$ ). Las diferencias observadas en este trabajo podrían estar relacionadas con las diferencias en dificultad de los distintos ítems. En todo caso, hay que tener también en cuenta de que se trata de una muestra de personas voluntarias altamente familiarizadas con la ciencia. Es de prever que el efecto de la dificultad sea mayor en la población general. Es algo que conviene estudiar en futuros trabajos.

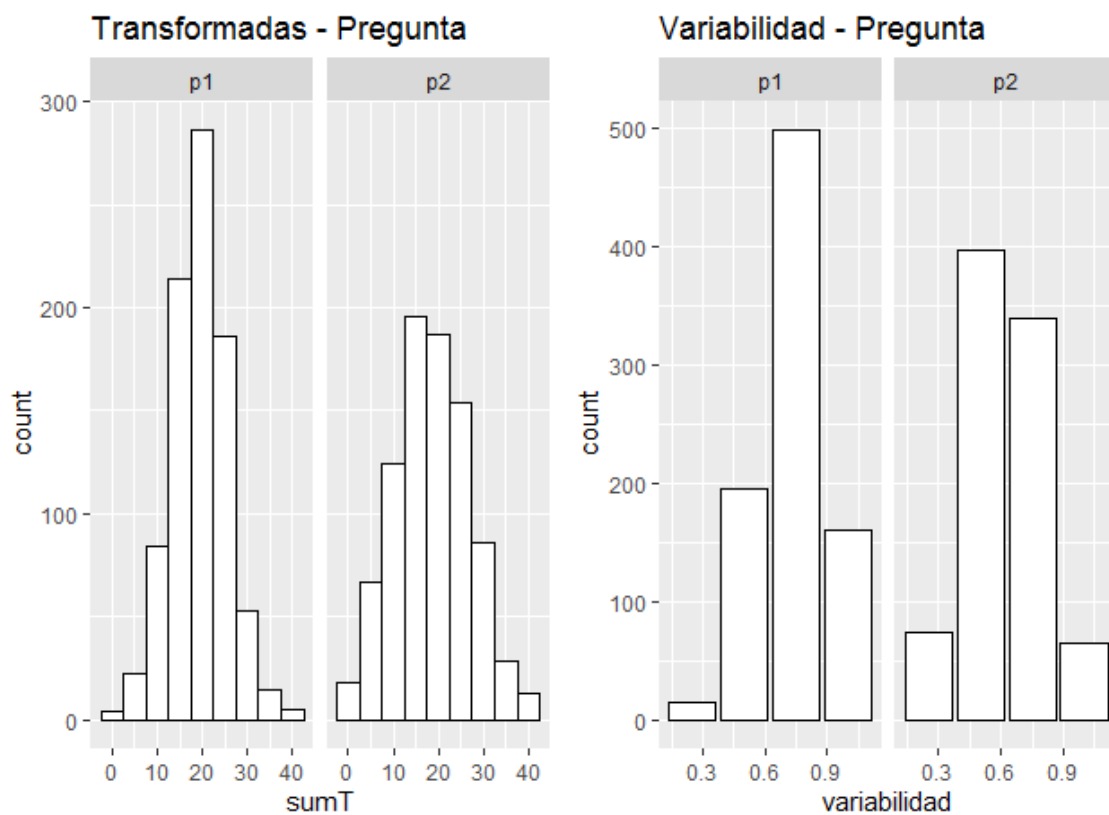
Se han encontrado pocas diferencias debidas al formato. De nuevo, las características de los participantes ya señaladas no permiten concluir si este resultado se debe a la escasa influencia real del formato de respuesta, o es atribuible, al menos parcialmente, a esas características. Es otra cuestión que necesita ser esclarecida en próximos estudios.

El análisis gráfico de las dos variables dependientes nos muestra que hay diferencias en las dos preguntas. Por lo que respecta la variable “sumT”, el histograma muestra cómo la distribución es más apuntada en la pregunta 1, destacando especialmente la mayor frecuencia de valores en el punto medio de la distribución. Algo similar ocurre con la variable “variabilidad”; en concreto, el diagrama de barras muestra que en la pregunta 1, el valor 0,75, que se corresponde con haber seleccionado tres respuestas distintas en los 4 ítems de cada pregunta, está presente en más de la mitad de las personas que han participado en el experimento; en cambio, en la pregunta 2, es un poco más frecuente el valor 0,5 (dos opciones de respuesta diferentes), es notablemente menor la frecuencia del valor 1 (haber seleccionado una categoría de respuesta distinta en cada ítem) y, en cambio, es mayor el número de participantes que han utilizado la misma categoría al responder a los cuatro ítems.

Por lo que respecta al análisis de las diferencias entre ambas preguntas, se observa que la media de sumT es igual a 20 en la pregunta 1, y a 19,12 en la pregunta 2. Las

desviaciones típicas son 6,14 y 8,21 respectivamente, es decir, hay más dispersión en la pregunta 2. En consecuencia, se incumple el supuesto de homocedasticidad y, por tanto, para compararlas se ha utilizado la prueba de Yuen. A partir de los resultados obtenidos, se rechaza la hipótesis nula de que no hay diferencias en las puntuaciones de ambas preguntas y se puede concluir que las puntuaciones en la pregunta 2 son más bajas que en la pregunta 1. Puesto que se trata de preguntas de A/D, se puede afirmar que las personas que han participado en el experimento tienden a estar más de acuerdo con las afirmaciones incluidas en la pregunta 1 que en las de la pregunta 2. El tamaño del efecto es medio-alto (0.46).

Figura 3 – Histograma y diagrama de barras de las variables dependientes



Por lo que respecta a la variabilidad, tal y como indica la figura 3, hay más variabilidad en la pregunta 1 - media y desviación típica son 0,73(0,17) y 0,61(0,19) respectivamente-. La comprobación de los supuestos para analizar las diferencias entre las preguntas nos lleva a rechazar tanto el de normalidad como el de homocedasticidad, por lo que aplicamos la prueba de Yuen. El resultado es estadísticamente significativo y el tamaño del efecto es medio-alto (0.43). El código utilizado se recoge en el Anexo VII.

### 3. CONCLUSIONES

Los resultados del experimento realizado indican que el efecto del formato de respuesta afecta de manera diferente a las dos variables dependientes (sumT y variabilidad) y lo hace también de manera diferente en función del contenido de la pregunta. Los cambios en la variable “sumT” no son deseables desde el punto de vista de la calidad de la medición; esta variable recoge las puntuaciones brutas y, por tanto, modificaciones debidas al formato de respuesta en esta variable estarían indicando errores sistemáticos de medición que generarían cambios en las conclusiones sobre las actitudes hacia el tema por el que se pregunte en la población de estudio. Es importante, no obstante, tener en cuenta las diferencias encontradas en función del contenido de la pregunta. El efecto es el contrario para la P1, de menor dificultad, que para la P2 (valorada como más difícil por los participantes). Este resultado parece estar indicando diferencias en el esfuerzo cognitivo implicado en la generación de una respuesta en función del contenido de la pregunta y el formato en el que se presenta la respuesta. Parecería que al responder a la pregunta 1 los participantes en el estudio han generado la respuesta positiva rápidamente, sin prestar excesiva atención y por eso se observa una mayor sensibilidad a la presencia de etiquetas (que requieren menos esfuerzo cognitivo) y al orden ad, en el que lo primero que se les presenta es el acuerdo. Es decir, parece razonable asumir que, una vez leída la pregunta, han generado una respuesta rápidamente y han buscado la categoría de respuesta que mejor representaba esa opinión. En orden ad, esas opciones estaban primero y, por tanto, han tendido a quedarse rápidamente con las opciones más positivas (las primeras). En el caso de recibir una categoría de respuesta sin etiqueta y/o en orden decreciente, el máximo acuerdo requería esforzarse más, dedicar más tiempo y, por tanto, no han llegado a procesar de manera completa las opciones de la escala de respuesta. Por el contrario, en la pregunta 2, más difícil, los resultados parecen indicar que las personas han dedicado mayor atención a procesar las opciones de la escala de respuesta y se podría decir, por tanto, que han ofrecido respuestas de mayor calidad. En todo caso, el efecto del orden de respuesta va en la línea de lo encontrado por Van Vaerenberh y Thomas (2012).

Desde el punto de vista de valorar la calidad de una encuesta, el incremento en la variabilidad es positivo, pues indica una mayor sensibilidad de las preguntas como herramientas de medición de las actitudes. En este caso no se han encontrado diferencias en la dirección del efecto debidas a la pregunta, sino diferencias en el efecto de las características del formato de respuesta. Es decir, el único factor que ha generado diferencias en la variabilidad de las respuestas a la pregunta 1 ha sido el

número de opciones. En cambio, en la pregunta 2 se han encontrado diferencias debidas al número de categorías, el número de opciones y las respuestas al problema del TRC. Estas diferencias indican, de nuevo, que la pregunta 2 ha sido más difícil de responder. Indican, también, que cuanto más difíciles de responder, mayor variabilidad en las respuestas.

Con independencia de que la mayor parte de los estudios publicados consideran que las escalas con siete categorías de respuesta son las que mejores resultados ofrecen, la realidad muestra que las encuestas de opinión utilizan, de manera casi exclusiva, una escala bipolar de cinco categorías con un punto intermedio neutral. Sin embargo, es probable que esta escala no esté contribuyendo a medir las opiniones y actitudes de manera precisa, pues solo deja disponibles dos categorías para definir la intensidad de la actitud. Si además tenemos en cuenta la influencia del estilo de respuesta, es probable que muchas personas solo dispongan de la posibilidad real de decir si están a favor, en contra o si no tienen una opinión formada al respecto (eligiendo el punto neutral). Y, en ese sentido, podría ser ajustada la recomendación de Peabody (1962) y Komorita (1963). Es algo que conviene explorar porque no hay duda de que los ítems bipolares son significativamente más sencillos de diseñar, de administrar y generan menos carga cognitiva en las personas encuestadas. Los resultados obtenidos en este estudio parecen ir en contra del supuesto de que para obtener respuestas de calidad hay que presentar a los sujetos preguntas sencillas, que puedan responder fácilmente y que no les generen carga cognitiva. Es evidente que esas características harán más fácil reclutar participantes, pero a costa de perder calidad en las respuestas. En todo caso, los resultados obtenidos apuntan a la necesidad ya señalada de prestar más atención a las preferencias de quienes tienen que responder a las encuestas, en lugar de dar por hecho qué es lo que quieren y necesitan. Como han señalado Preston y Coleman (2000), ya que tienen que tomarse el esfuerzo de responder a una encuesta, los encuestados prefieren enfrentarse a una que les permita expresar adecuadamente su opinión. Esto enlaza también con la importancia atribuida por algunos autores a la motivación (p.e., Krosnick, 1991; Schwarz y Vaughn, 2002).

En todo caso, esta es una conclusión muy preliminar, pues este estudio presenta importantes limitaciones que deberán irse abordando en futuras investigaciones. En concreto, se han utilizado muy pocos ítems, hay muchos otros formatos que deben ser explorados, sería conveniente poder comparar de manera completa las dos preguntas y, para ello, los participantes deberían recibir el mismo formato para responder a ambas y, por último, es necesario replicar el estudio en la población general, ya que los resultados pueden ser notablemente diferentes.

Para concluir, el principal resultado de este estudio va en la línea de la mayor parte de los estudios publicados al respecto al mostrar que la calidad de las respuestas, entendida como mayor precisión, aumenta cuando lo hace el número de opciones de la escala de respuesta. Por otro lado, si realmente se busca hacer realidad el supuesto de que la actividad de responder a una encuesta está dirigida por el sistema 2 de procesamiento, parece evidente que debemos contribuir a generar opciones para que ese sistema tome el control sobre el automático, rápido y asociativo. Parece que el diseño de las preguntas (contenido y escala de respuesta) son elementos clave para contribuir a alcanzar ese objetivo.

## BIBLIOGRAFÍA

Allum, N., Sturgis, P., Tabourazi, D. y Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: a meta-analysis. *Public Understanding of Science*, 17 (1), 35-54.

Alwin, D. F. (1992). Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. *Sociological Methodology*, 22, 83-118.

Alwin, D. F y Krosnick, J. A. (1991). The reliability of survey attitude measurement. The influence of question and respondent attributes. *Sociological Methods & Research*, 20 (1), 139-181.

Andrews, F. M. (1984). Construct validity and error components of survey measures: a structural modeling approach. *Public Opinion Quarterly*, 48(2), 409–442.

Bendig, A. W. (1953). The reliability of self-ratings as a function of the amount of verbal anchoring and the number of categories on the scale. *The Journal of Applied Psychology*, 37(1), 38-41.

Bendig, A. W. y Hughes, J. B. (1953). Effect of amount of verbal anchoring and number of rating-scale categories upon transmitted information. *Journal of Experimental Psychology*, 46(2), 87-90.

Blagotic, A. y Daróczi, G. (2014). Rapport: a report templating system. R package version 1.0, en: <http://cran.r-project.org/package=rapport>.

Dillman, D. A. et al. (1995). Effects of category order on answers in mail and telephone surveys. *Rural Sociology*, 60(4), 674-687.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.

Garner, W. R. (1960). Rating scales, discriminability, and information transmission. *The Psychological Review*, 67(6), 343-352.

Hipkins, R., Stockwell, W. W., Bolstad, R., y Baker, R. (2002). *Commonsense, Trust, and Science. How Patterns of Beliefs and Attitudes to Science Pose Challenges for Effective Communication*. New Zealand: Ministry of Research, Science, and Technology. Disponible en:

<https://www.nzcer.org.nz/research/publications/commonsense-trust-and-science-how-patterns-beliefs-and-attitudes-science-pose->, consultado el 29 de mayo de 2018.

Kalton, G., Collins, M. y Brook, L. (1978). Experiments in wording opinion questions. *Applied Statistics*, 27(2), 149-161.

Kalton, G., y Schuman, H. (1982). The Effect of the Question on Survey Responses: A Review. *Journal of the Royal Statistical Society. Series A (General)*, 145(1), 42-73.

Kahneman, D. (2011): *Pensar rápido, pensar despacio*, Debate, Barcelona.

Komorita, S. S. (1963). Attitude content, intensity, and the neutral point on a Likert scale. *The Journal of Social Psychology*, 61, 327-334.

Komorita, S. S. y Graham, W. K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement*, 25(4), 987-995.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J. A. y Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51, 201-219.

Krosnick, J. A., Narayan, S. y Smith, W. R. (1996). Satisficing in surveys: initial evidence. En M. T. Baverman y J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29-44). San Francisco: Jossey-Bass.

Lerner, J.S. y Tetlock, P.E. (1999) Accounting for the Effects of Accountability. *Psychological Bulletin*, 125, 255-275.

Leung, S-O (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, 37, 412-421.

Leyens, J. P y Codol, J. P. (1992). Cognición social. En M. Hewstone, W. Stroebe, J.P. Codol y G. M. Stephenson (Coords.), *Introducción a la Psicología Social. Una perspectiva europea* (pp. 99-119). Barcelona: Editorial Ariel.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 140-155.

Mair, P., & Wilcox, R. R. (2019). Robust Statistical Methods in R Using the WRS2 Package. *Behavior Research Methods* <https://doi.org/10.3758/s13428-019-01246-w>.

Mattel, M. S. y Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? *Journal of Applied Psychology*, 56(6), 506-509.

McKelvie, S. J. (1978). Graphic rating scales- How many categories? *British Journal of Psychology*, 69, 185-202.

Muñiz, J. (Coord) (1996). *Psicometría*. Madrid: Editorial Universitas.

Muñoz van den Eynde, A. (2014). Reflexión cognitiva. Implicaciones para la validez de las encuestas de percepción social de la ciencia y la tecnología. En A. Muñoz van den Eynde y E. H. Lopera Pareja (coords.), *La Percepción Social de la Ciencia. Claves para la Cultura Científica* (pp. 47-79). Madrid: Los Libros de la Catarata.

Peabody, D. (1962). Two components in bipolar scales: Direction and extremeness. *Psychological Review*, 69(2), 65-73.

Pratkanis, A. R. y Greenwald, A. G. (1989). A Sociocognitive Model of Attitude Structure and Function. *Advances in Experimental Social Psychology*, 22, 245-285.

Preston, C. C. y Colman, M. A. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, en: <https://www.R-project.org/>.

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, en: <https://CRAN.R-project.org/package=psych>. Version = 1.8.12.

Revilla, M. A., Saris, W. E. y Krosnick, J. A. (2013). Choosing the number of categories in Agree-Disagree scales , 43(1), 73-97.

Saris, W. E., Revilla, M., Krosnick, J. A., y Shaeffer, E. M. (2010). Comparing questions with agree/disagree response option to questions with item-specific response options. *Survey Research Methods*, 4, 61-79.

Schuman, H. y Presser, S. (1996). *Questions & Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context*. San Diego: Sage Publications.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21, 277-87.

Schwarz, N. y Vaughn, A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. En T. Gilovich, D. Griffin y D. Kahnemann (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 103-119). New York: Cambridge University Press.

Searle, J. R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Shoemaker, P. J., Eichholz, M., y Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, 14(2), 193-201.

Sturgis, P. y Smith, P. (2010). Fictitious issues revisited: political interest, knowledge and the generation of non-attitudes. *Political Studies*, 58, 66-84.

Sturgis, P., Roberts, C. y Smith, P. (2014). Middle alternatives revisited: How the neither/nor response act as a way of saying 'I don't know'. *Sociological Methods & Research*, 43, 15-38.

Toepoel, V., Das, M. y Van Soest, A. (2008). Effects of design in web surveys. Comparing trained and fresh respondents. *Public Opinion Quarterly*, 72(5), 985-1007.

Toplak, M. E., West, R. F. y Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 147-168. [El ítem de TRC incluido en el experimento].

Tourangeau, R., Rips, L. J. y Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Van der Pligt, J., Eiser, J. R. y Spears, R. (1987). Comparative judgment and preference: the influence of the number of response alternatives. *British Journal of Social Psychology*, 26, 269-280.

Van Vaerenbergh, Y. y Thomas, T. D. (2012). Response styles in survey research: A literature review of antecedents, consequences and remedies. *International Journal of Public Opinion Research*, 25(3), 195-217.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wyatt, R. C. y Meyers, L. S. (1987). Psychometric properties of four 5-point Likert-type response scales. *Educational and Psychological Measurement*, 47(1), 27-35.



## ANEXO I - Diseño del experimento

Pregunta 1

Opción 1

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones?

	Totalmente de acuerdo	Bastante de acuerdo	Ni de acuerdo ni en desacuerdo	Bastante en desacuerdo	Totalmente en desacuerdo
Las asignaturas de ciencia siempre se me dieron mal					
La ciencia es tan especializada que me cuesta entenderla					
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana					
La información que oigo acerca de la ciencia suele ser cierta					

Esto está definido en el fichero “excope” como: p1\_o1\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o1 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parImpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “1”, etiqueta “totalmente” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 2

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones?

	<b>Muy de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Ni de acuerdo ni en desacuerdo</b>	<b>Bastante en desacuerdo</b>	<b>Muy en desacuerdo</b>
Las asignaturas de ciencia siempre se me dieron mal					
La ciencia es tan especializada que me cuesta entenderla					
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana					
La información que oigo acerca de la ciencia suele ser cierta					

Esto está definido en el fichero “excope” como: p1\_o2\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o2 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 3

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones?

	<b>Muy de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Poco de acuerdo</b>	<b>Nada de acuerdo</b>
Las asignaturas de ciencia siempre se me dieron mal				
La ciencia es tan especializada que me cuesta entenderla				
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana				
La información que oigo acerca de la ciencia suele ser cierta				

Esto está definido en el fichero “excope” como: p1o3\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o3 significa:

- 1) Valor “4” en la variable “opciones”.
- 2) Valor “1”, etiqueta “par” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P1 implica:

Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 4

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones?

	<b>Muy en desacuerdo</b>	<b>Bastante en desacuerdo</b>	<b>Algo en desacuerdo</b>	<b>Algo de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Muy de acuerdo</b>
Las asignaturas de ciencia siempre se me dieron mal						
La ciencia es tan especializada que me cuesta entenderla						
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana						
La información que oigo acerca de la ciencia suele ser cierta						

Esto está definido en el fichero “excope” como: p1\_o4\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o4 significa:

- 1) Valor “6” en la variable “opciones”.
- 2) Valor “1”, etiqueta “par” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 5

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones? Dispones de una escala de 1 a 5 en la que el 1 significa que no estás nada de acuerdo y el 5 significa que estás muy de acuerdo.

	1	2	3	4	5
Las asignaturas de ciencia siempre se me dieron mal					
La ciencia es tan especializada que me cuesta entenderla					
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana					
La información que oigo acerca de la ciencia suele ser cierta					

Esto está definido en el fichero “excope” como: p1\_o5\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o5 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “2”, etiqueta “no\_etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 6

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones?

	<b>Muy de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Algo de acuerdo</b>	<b>Poco de acuerdo</b>	<b>Nada de acuerdo</b>
Las asignaturas de ciencia siempre se me dieron mal					
La ciencia es tan especializada que me cuesta entenderla					
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana					
La información que oigo acerca de la ciencia suele ser cierta					

Esto está definido en el fichero “excope” como: p1\_o6\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o6 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 7

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones? Dispones de una escala de 1 a 4 en la que el 1 significa que no estás nada de acuerdo y el 4 significa que estás muy de acuerdo.

	1	2	3	4
Las asignaturas de ciencia siempre se me dieron mal				
La ciencia es tan especializada que me cuesta entenderla				
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana				
La información que oigo acerca de la ciencia suele ser cierta				

Esto está definido en el fichero “excope” como: p1\_o7\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o7 significa:

- 1) Valor “4” en la variable “opciones”.
- 2) Valor “1”, etiqueta “par” en la variable “parimpar”
- 3) Valor “2”, etiqueta “no\_etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 1

### Opción 8

En diferentes estudios se ha venido recabando la opinión de los ciudadanos sobre la ciencia, los científicos o cuestiones relacionadas con el control de su actividad. A continuación, te presentamos algunas de las respuestas proporcionadas por otras personas en estos estudios. ¿Hasta qué punto estás de acuerdo con sus afirmaciones?

	<b>Totalmente en desacuerdo</b>	<b>Bastante en desacuerdo</b>	<b>Algo en desacuerdo</b>	<b>Algo de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Totalmente de acuerdo</b>
Las asignaturas de ciencia siempre se me dieron mal						
La ciencia es tan especializada que me cuesta entenderla						
Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana						
La información que oigo acerca de la ciencia suele ser cierta						

Esto está definido en el fichero “excope” como: p1\_o8\_:

a, para: Las asignaturas de ciencia siempre se me dieron mal

b, para: La ciencia es tan especializada que me cuesta entenderla

c, para: Considero que es necesario saber sobre ciencia y tecnología en mi vida cotidiana

d, para: La información que oigo acerca de la ciencia suele ser cierta

Por tanto, o8 significa:

- 1) Valor “6” en la variable “opciones”.
- 2) Valor “1”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “1”, etiqueta “totalmente” en la variable “extremo”

Además, P1 implica: Valor “1”, etiqueta “p1” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones?

### Opción 1

	<b>Totalmente de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Ni de acuerdo ni en desacuerdo</b>	<b>Bastante en desacuerdo</b>	<b>Totalmente en desacuerdo</b>
La investigación científica depende cada vez más de intereses comerciales					
El desarrollo científico y tecnológico puede tener consecuencias indeseadas					
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos					
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos					

Esto está definido en el fichero “excope” como: p2\_o1\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o1 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “1”, etiqueta “totalmente” en la variable “extremo”
- 6) Valor “12, etiqueta “ad” en la variable “orden”

Además, P2 implica:

Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones?

### Opción 2

	<b>Muy de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Ni de acuerdo ni en desacuerdo</b>	<b>Bastante en desacuerdo</b>	<b>Muy en desacuerdo</b>
La investigación científica depende cada vez más de intereses comerciales					
El desarrollo científico y tecnológico puede tener consecuencias indeseadas					
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos					
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos					

Esto está definido en el fichero “excope” como: p2\_o2\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o2 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P2 implica: Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones?

### Opción 3

	<b>Muy de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Poco de acuerdo</b>	<b>Nada de acuerdo</b>
La investigación científica depende cada vez más de intereses comerciales				
El desarrollo científico y tecnológico puede tener consecuencias indeseadas				
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos				
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos				

Esto está definido en el fichero “excope” como: p2\_o3\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o3 significa:

- 1) Valor “4” en la variable “opciones”.
- 2) Valor “1”, etiqueta “par” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P2 implica:

Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones?

### Opción 4

	Muy en desacuerdo	Bastante en desacuerdo	Algo en desacuerdo	Algo de acuerdo	Bastante de acuerdo	Muy de acuerdo
La investigación científica depende cada vez más de intereses comerciales						
El desarrollo científico y tecnológico puede tener consecuencias indeseadas						
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos						
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos						

Esto está definido en el fichero “excope” como: p2\_o4\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o4 significa:

- 1) Valor “6” en la variable “opciones”.
- 2) Valor “1”, etiqueta “par” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P2 implica: Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones? Dispones de una escala de 1 a 4 en la que el 1 significa que no estás nada de acuerdo y el 5 significa que estás muy de acuerdo.

### Opción 5

	1	2	3	4	5
La investigación científica depende cada vez más de intereses comerciales					
El desarrollo científico y tecnológico puede tener consecuencias indeseadas					
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos					
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos					

Esto está definido en el fichero “excope” como: p2\_o5\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o5 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “2”, etiqueta “no\_etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P2 implica: Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones?

### Opción 6

	<b>Muy de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Algo de acuerdo</b>	<b>Poco de acuerdo</b>	<b>Nada de acuerdo</b>
La investigación científica depende cada vez más de intereses comerciales					
El desarrollo científico y tecnológico puede tener consecuencias indeseadas					
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos					
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos					

Esto está definido en el fichero “excope” como: p2\_o6\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o6 significa:

- 1) Valor “5” en la variable “opciones”.
- 2) Valor “2”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “1”, etiqueta “ad” en la variable “orden”

Además, P2 implica: Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones? Dispones de una escala de 1 a 4 en la que el 1 significa que no estás nada de acuerdo y el 4 significa que estás muy de acuerdo.

### Opción 7

	1	2	3	4
La investigación científica depende cada vez más de intereses comerciales				
El desarrollo científico y tecnológico puede tener consecuencias indeseadas				
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos				
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos				

Esto está definido en el fichero “excope” como: p2\_o7\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o7 significa:

- 1) Valor “4” en la variable “opciones”.
- 2) Valor “1”, etiqueta “par” en la variable “parimpar”
- 3) Valor “2”, etiqueta “no\_etiquetado” en la variable “etiqueta”
- 4) Valor “1”, etiqueta “unipolar” en la variable “polaridad”
- 5) Valor “2”, etiqueta “muy” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P2 implica: Valor “2”, etiqueta “p2” en la variable “pregunta”

## Pregunta 2

¿Y hasta qué punto estás de acuerdo con estas afirmaciones?

### Opción 8

	<b>Totalmente en desacuerdo</b>	<b>Bastante en desacuerdo</b>	<b>Algo en desacuerdo</b>	<b>Algo de acuerdo</b>	<b>Bastante de acuerdo</b>	<b>Totalmente de acuerdo</b>
La investigación científica depende cada vez más de intereses comerciales						
El desarrollo científico y tecnológico puede tener consecuencias indeseadas						
Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos						
Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos						

Esto está definido en el fichero “excope” como: p2\_o8\_:

a, para: La investigación científica depende cada vez más de intereses comerciales

b, para: El desarrollo científico y tecnológico puede tener consecuencias indeseadas

c, para: Los avances científicos se producen con tanta rapidez que es difícil que las autoridades puedan controlarlos

d, para: Es poco probable que los ciudadanos lleguemos a saber realmente cuáles son los riesgos de los avances científicos

Por tanto, o8 significa:

- 1) Valor “6” en la variable “opciones”.
- 2) Valor “1”, etiqueta “impar” en la variable “parimpar”
- 3) Valor “1”, etiqueta “etiquetado” en la variable “etiqueta”
- 4) Valor “2”, etiqueta “bipolar” en la variable “polaridad”
- 5) Valor “1”, etiqueta “totalmente” en la variable “extremo”
- 6) Valor “2”, etiqueta “da” en la variable “orden”

Además, P2 implica: Valor “2”, etiqueta “p2” en la variable “pregunta”

Pregunta 3:

Problema

Si el laboratorio A tarda 6 días en vaciar una bombona de nitrógeno líquido y el laboratorio B tarda 12 días, ¿cuántos días tardarían los dos laboratorios en vaciar una bombona de nitrógeno líquido si tuvieran que compartirla?

a) 3

b) 4

c) 8

d) 9

En el fichero aparece como: "problema" con tres valores: 0, 1 y 2:

0 = Incorrecto

1 = Correcto

2 = Abandono

## ANEXO II - Código para el ajuste del fichero de datos

```
# Cargamos los datos
```

```
load("J:/Privada/Ana/Máster Estadística con R/TFM/Código/excope.RData")
```

```
# Limpiamos los nombres para trabajar mejor el tratamiento de datos
```

```
# Vector P1-P2
```

```
pregunta <- paste0("p", 1:2)
```

```
# Vector o1-o8
```

```
opcion <- paste0("o", 1:8)
```

```
# Vector a-d
```

```
letra <- letters[1:4]
```

```
# Vector guión bajo
```

```
guion_bajo <- "_ "
```

```
# Data Frame con todas las posibles combinaciones
```

```
nombres_df <- expand.grid(pregunta, guion_bajo, opcion, guion_bajo, letra)
```

```
# Lo ordenamos por pregunta y observación
```

```
nombres_ordenados <- nombres_df[order(nombres_df$Var1, nombres_df$Var3),]
```

```
# Los unimos obteniendo un vector
```

```
nombres_vector <- do.call(paste0, nombres_ordenados)
```

```
# añadimos manualmente muestra y problema a lo anterior
```

```
nombres_correctos <- c("muestra", nombres_vector, "problema")
```

```
# Tenemos los nombres
```

```
nombres_correctos
```

```
# Asignamos los nombres correctos
```

```
names(excope) <- nombres_correctos
```

```
names(excope)
```

```
# Pasamos a formato largo para tener todas las opciones en una única variable:
```

```
excope_l <- reshape(excope,
```

```
  varying = 2:65,
```

```
  v.names = "valores",
```

```
  timevar = "variables",
```

```
  times = names(excope[2:65]),
```

```
  direction = "long")
```

```
#Eliminamos los NA:
```

```
excope_total <- excope_l[complete.cases(excope_l),]
```

```
str(excope_total)
```

```
# Desarrollamos una función para definir las variables independientes y crear las condiciones para genera las variables dependientes
```

```
tratar <- function(x) {
```

```
  datos_separados <- strsplit(x, split = "_")
```

```
  p <- sapply(datos_separados, "[", 1)
```

```
  pregunta <- ifelse(p == "p1", 1, 2)
```

```
  abcd <- sapply(datos_separados, "[", 3)
```

```

# pregunta <- factor(pregunta)
q <- sapply(datos_separados, "[", 2)
opciones <- ifelse((q == "o3" | q == "o7"), 4,
  ifelse((q == "o1" | q == "o2" | q == "o5" | q == "o6"), 5, 6))
# opciones <- factor(opciones, labels = c(4, 5, 6))
parimpar <- ifelse((q == "o1" | q == "o2" | q == "o5" | q == "o6"), 1, 2)
parimpar <- factor(parimpar, labels = c("impar", "par"))
etiqueta <- ifelse((q == "o5" | q == "o7"), 2, 1)
etiqueta <- factor(etiqueta, labels = c("etiquetado", "no_etiquetado"))
polaridad <- ifelse((q == "o3" | q == "o5" | q == "o6" | q == "o7"), 1, 2)
polaridad <- factor(polaridad, labels = c("unipolar", "bipolar"))
extremo <- ifelse((q == "o1" | q == "o8"), 1, 2)
extremo <- factor(extremo, labels = c("totalmente", "muy"))
orden <- ifelse((q == "o1" | q == "o2" | q == "o3" | q == "o6"), 1, 2)
orden <- factor(orden, labels = c("ad", "da"))

return(data.frame(pregunta, abcd, opciones, parimpar, etiqueta, polaridad, extremo,
orden))
}

# Tratamos los datos
encuesta <- tratar(excope_total$variables)
str(encuesta)

encuesta$id <- excope_total$id # incluimos id en nuevos datos
encuesta$valores <- excope_total$valores # incluimos valores respuesta en nuevos
datos
encuesta$muestra <- factor(excope_total$muestra, labels = c("general", "Ciemat")) #
incluimos la variable muestra en los nuevos datos
encuesta$problema <- factor(excope_total$problema, labels = c("incorrecto",
"correcto", "abandono")) # incluimos la variable problema en los nuevos datos

encuesta <- encuesta[order(encuesta$id),] # ordenamos por identidad
head(encuesta)

# Creamos una variable con una identidad que nos sirva para calcular las variables
dependientes
encuesta$id_tmp <- paste0(as.character(encuesta$id), "_",
as.character(encuesta$pregunta), sep = "")
str(encuesta)

# Calculamos el índice de variabilidad
variabilidad <- vector("numeric", dim(encuesta)[1])
for (i in 1:dim(encuesta)[1]) {
  identidad <- encuesta[i, "id_tmp"]
  variabilidad[i] <- length(unique(encuesta$valores[encuesta$id_tmp == identidad])) / 4
}
encuesta$variabilidad <- variabilidad

```

```
# Creamos las puntuaciones transformadas:
```

```
transformadas <- vector("numeric", dim(encuesta)[1])
for (i in 1:dim(encuesta)[1]) {
  if(encuesta$opciones[i] == 4) {
    transformadas[i] <- ((encuesta$valores[i] - 1)/3) * 10
  } else if(encuesta$opciones[i] == 5) {
    transformadas[i] <- ((encuesta$valores[i] - 1)/4) * 10
  } else {
    transformadas[i] <- ((encuesta$valores[i] - 1)/5) * 10
  }
}
encuesta$transformadas <- transformadas
```

```
#ordenamos columnas
```

```
encuesta <- encuesta[, c("id", "id_tmp", "abcd", "pregunta",
  "opciones", "parimpar", "etiqueta", "polaridad", "extremo",
  "orden", "muestra", "problema", "valores", "variabilidad",
  "transformadas")]
```

```
# cambiamos nombre filas.
```

```
row.names(encuesta) <- 1:dim(encuesta)[1]
```

```
# Observamos resultados
```

```
head(encuesta, 20)
```

```
str(encuesta)
```

```
# Transformamos "pregunta" y "opciones" en factores:
```

```
encuesta$pregunta <- factor(encuesta$pregunta, labels = c("p1", "p2"))
```

```
encuesta$opciones <- factor(encuesta$opciones, labels = c("4", "5", "6"))
```

#Hace falta devolver el fichero al formato ancho, para que haya una observación por sujeto. Hay que recordar que la variable "variabilidad" contabiliza el número de opciones diferentes entre p1\_a, p1\_b, p1\_c y p1\_d para la pregunta 1 y para cada sujeto. Y entre p2\_a, p2\_b, p2\_c y P2\_d para la pregunta 2 y para cada sujeto.

```
# Formato ancho:
```

```
encuesta_ancho <- reshape(encuesta,
  v.names = c("valores", "transformadas"),
  timevar = "abcd",
  idvar = "id_tmp",
  direction = "wide")
str(encuesta_ancho)
```

```

# Ordenamos dataset
encuesta_ancho <-
  encuesta_ancho[, c(names(encuesta_ancho[1:11]),
    "valores.a", "valores.b", "valores.c", "valores.d",
    "transformadas.a", "transformadas.b", "transformadas.c", "transformadas.d")]
head(encuesta_ancho)
str(encuesta_ancho)

# eliminamos variable id_tmp ya que no va a ser necesaria más adelante

encuesta_ancho$id_tmp <- NULL

#Añadimos una variable que sea la suma de las transformadas en los cuatro ítems

encuesta_ancho$sumT <- (encuesta_ancho$transformadas.a +
encuesta_ancho$transformadas.b +
  encuesta_ancho$transformadas.c + encuesta_ancho$transformadas.d)
summary(encuesta_ancho$sumT)

# Eliminamos los NA

excope_18 <- encuesta_ancho[complete.cases(encuesta_ancho),]

# Guardamos el fichero:

save(excope_18, file = "excope_18.RData")

# Fichero de datos separado por p1 y p2
excope_18_p1 <- excope_18[excope_18$pregunta == "p1",]
excope_18_p2 <- excope_18[excope_18$pregunta == "p2",]

```

### **ANEXO III – Análisis de la variable sumT en la pregunta 1**

```
# Variable PUNTUACIONES TRANSFORMADAS:
```

```
# Fichero de datos: excope_18
```

```
# Cargamos los datos
```

```
load("J:/Privada/Ana/Máster Estadística con R/TFM/Código/excope_18.RData")
```

```
# HIPÓTESIS:
```

```
# H0: La media de las puntuaciones transformadas es igual para los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
# H1: La media de las puntuaciones transformadas difiere entre los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
#1) PREGUNTA 1
```

```
# Creamos el fichero con los datos para la pregunta 1:
```

```
excope_18_p1 <- excope_18[excope_18$pregunta == "p1",]
```

```
# Análisis descriptivo VD:
```

```
summary(excope_18_p1$sumT)
```

```
sd(excope_18_p1$sumT)
```

```
hist(excope_18_p1$sumT,  
      main = "Histograma P1 - Suma P. Transformadas",  
      xlab = "Puntuaciones transformadas")
```

```
# VI: opciones
```

```
# Análisis descriptivo:
```

```
library(psych)
```

```
describeBy(excope_18_p1$sumT, excope_18_p1$opciones)
```

```
library(ggplot2)
```

```
ggplot(excope_18_p1, aes(factor(opciones), sumT)) +  
  geom_boxplot() +  
  labs(title = "Puntuaciones transformadas P1- Opciones")
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
# El gráfico de cajas indica que hay.
```

```
# Para identificar los valores:
```

```

library(rapportools)

rp.outlier(excope_18_p1[excope_18_p1$opciones == "4", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$opciones == "5", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$opciones == "6", "sumT"])

# La función rp.outlier no identifica valores extremos

# Comprobación del supuesto de normalidad

by(excope_18_p1$sumT, excope_18_p1$opciones, shapiro.test)

# Rechazamos la asunción del supuesto de normalidad para los tres grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando la prueba de Flinger-Killeen (no paramétrica)

fligner.test(excope_18_p1$sumT, excope_18_p1$opciones)

# Mantenemos el supuesto de homogeneidad de varianzas.

# Aunque se cumple el supuesto de homocedasticidad, no se cumple el de normalidad
y hay outliers, así que utilizaremos la prueba robusta.

library(WRS2)
t1way(sumT ~ opciones, excope_18_p1)

# Mantenemos H0. No hay evidencia de que haya diferencias en las puntuaciones de
los ítems de la pregunta 1 relacionadas con el número de opciones de la escala de
respuesta.

# VI: parimpar

# Análisis descriptivo:

describeBy(excope_18_p1$sumT, excope_18_p1$parimpar)
ggplot(excope_18_p1, aes(factor(parimpar), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1- Parimpar")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores

rp.outlier(excope_18_p1[excope_18_p1$parimpar == "par", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$parimpar == "impar", "sumT"])

# Constatamos que no hay.

```

```

# Comprobación del supuesto de normalidad

by(excope_18_p1$sumT, excope_18_p1$parimpar, shapiro.test)

# Rechazamos la asunción del supuesto de normalidad para los dos grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Flinger-Killeen:

fligner.test(excope_18_p1$sumT, excope_18_p1$parimpar)

# La evidencia obtenida nos lleva a asumir el supuesto de homocedasticidad, pero
como hay outlier y no se cumple el de normalidad, utilizaremos la prueba robusta

yuen(sumT ~ parimpar, excope_18_p1)

# Rechazamos la hipótesis nula, pero es un efecto marginal, así que no consideramos
que el resultado sea relevante.

#VI: etiqueta

# Análisis descriptivo:

describeBy(excope_18_p1$sumT, excope_18_p1$etiqueta)
ggplot(excope_18_p1, aes(factor(etiqueta), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1 - Etiqueta")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que los hay. Comprobamos para tratar de identificarlos.

rp.outlier(excope_18_p1[excope_18_p1$etiqueta == "etiquetado", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$etiqueta == "no_etiquetado", "sumT"])

# En este caso la función sí lo identifica, en el valor 40. Se trataría de un caso de estilo
extremo.

# Comprobación del supuesto de normalidad

by(excope_18_p1$sumT, excope_18_p1$etiqueta, shapiro.test)

# No se cumple el supuesto de normalidad, por lo que comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Flinger-Killeen:

fligner.test(excope_18_p1$sumT, excope_18_p1$etiqueta)

```

# Mantenemos el supuesto de homocedasticidad, pero como hemos encontrado outliers y no se cumple el de normalidad, utilizamos la prueba robusta

```
yuen(sumT ~ etiqueta, excope_18_p1)
```

# Rechazamos la H0 y, por tanto, hemos encontrado evidencia de que hay diferencias en la puntuación a los ítems de la p1 debidas a la presencia, o no, de etiquetas. La media es significativamente menor cuando no se incluyen etiquetas. No obstante, hay que tener en cuenta que, tal y como se han diseñado las opciones en el experimento, hay coincidencia entre no tener etiqueta e ir en la dirección da en lugar de ad. El tamaño del efecto es moderado (0.31).

# VI: polaridad

# Análisis descriptivo:

```
describeBy(excope_18_p1$sumT, excope_18_p1$polaridad)
ggplot(excope_18_p1, aes(factor(polaridad), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1 - Polaridad")
```

# Comprobación de los supuestos:

# Identificación de outliers:

# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores

```
rp.outlier(excope_18_p1[excope_18_p1$polaridad == "unipolar", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$polaridad == "bipolar", "sumT"])
```

# En este caso la función no los identifica

# Comprobación del supuesto de normalidad

```
by(excope_18_p1$sumT, excope_18_p1$polaridad, shapiro.test)
```

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Flinger-Killeen:

```
fligner.test(excope_18_p1$sumT, excope_18_p1$polaridad)
```

# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad, además de que el gráfico de cajas identifica algunos outliers. Por tanto, utilizaremos la prueba robusta:

```
yuen(sumT ~ polaridad, excope_18_p1)
```

# En este caso, rechazamos H0, pero el nivel de significación es marginal. La medida del tamaño del efecto (0.1) nos indica que no es una diferencia relevante.

```
# VI: extremo
```

```
# Análisis descriptivo:
```

```
describeBy(excope_18_p1$sumT, excope_18_p1$extremo)
ggplot(excope_18_p1, aes(factor(extremo), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1 - Extremo")
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
# El gráfico de cajas indica que los hay. Comprobamos para tratar de identificar los valores.
```

```
rp.outlier(excope_18_p1[excope_18_p1$extremo == "totalmente", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$extremo == "muy", "sumT"])
```

```
# La función no los identifica
```

```
# Comprobación del supuesto de normalidad
```

```
by(excope_18_p1$sumT, excope_18_p1$extremo, shapiro.test)
```

```
# No se cumple el supuesto de normalidad en los dos grupos, por lo que hay que rechazarlo. Comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:
```

```
fligner.test(excope_18_p1$sumT, excope_18_p1$extremo)
```

```
# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad y se supone que hay outliers, por lo que usamos la prueba robusta:
```

```
yuen(sumT ~ extremo, excope_18_p1)
```

```
# Mantenemos H0. No hay evidencia de que las puntuaciones en los 4 ítems de la p1 varíen de manera significativa debido a la etiqueta utilizada para definir el extremo de la escala.
```

```
# VI: orden
```

```
# Análisis descriptivo:
```

```
describeBy(excope_18_p1$sumT, excope_18_p1$orden)
ggplot(excope_18_p1, aes(factor(orden), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1 - Orden")
```

```
# Comprobación de los supuestos:
```

```

# Identificación de outliers:
# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores

rp.outlier(excope_18_p1[excope_18_p1$orden == "ad", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$orden == "da", "sumT"])

# Comprobación del supuesto de normalidad

by(excope_18_p1$sumT, excope_18_p1$orden, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p1$sumT, excope_18_p1$orden)

# Mantenemos el supuesto de homocedasticidad, pero como parece haber outliers y
no se cumple el supuesto de normalidad, utilizaremos la prueba robusta:

yuen(sumT ~ orden, excope_18_p1)

# Rechazamos H0. Hay evidencia de que el orden en el que se describe la escala
influye en las puntuaciones medias. El tamaño del efecto es bajo. En todo caso, al
tratarse de escalas inversas, lo que el resultado nos está indicando es que la opinión
media de las personas que han participado en el estudio se sitúa algo por encima de la
puntuación media con ambos tipos de escala. Se ve muy bien en el gráfico de cajas.

# De todos modos, visto el resultado obtenido con la variable etiqueta y la variable
orden, conviene analizar el efecto conjunto de ambas variables.

# VI: Muestra

# Análisis descriptivo:

describeBy(excope_18_p1$sumT, excope_18_p1$muestra)
ggplot(excope_18_p1, aes(factor(muestra), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1 - Muestra")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que los hay. Comprobamos para identificarlos

rp.outlier(excope_18_p1[excope_18_p1$muestra == "general", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$muestra == "general", "sumT"])

# Comprobación del supuesto de normalidad

by(excope_18_p1$sumT, excope_18_p1$muestra, shapiro.test)

```

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```
fligner.test(excope_18_p1$sumT, excope_18_p1$muestra)
```

# Mantenemos el supuesto de homocedasticidad, pero como no se cumple el de normalidad y parece haber outliers, utilizamos la prueba robusta:

```
yuen(sumT ~ muestra, excope_18_p1)
```

# Mantenemos la H0. No hemos obtenido evidencia de que haya diferencias en las puntuaciones entre la muestra de la población general y la muestra del Ciemat. No debemos olvidar que se trata de un conjunto de voluntarios muy interesados por la ciencia. Por otro lado, en la muestra general se detectan más casos de estilo extremo de respuesta.

# VI: problema

# Análisis descriptivo:

```
describeBy(excope_18_p1$sumT, excope_18_p1$problema)
ggplot(excope_18_p1, aes(factor(problema), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P1 - Problema")
```

# Comprobación de los supuestos:

# Identificación de outliers:

# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores

```
rp.outlier(excope_18_p1[excope_18_p1$problema == "incorrecto", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$problema == "correcto", "sumT"])
rp.outlier(excope_18_p1[excope_18_p1$problema == "abandono", "sumT"])
```

# En este caso la función identifica un outlier en el valor 40.

# Comprobación del supuesto de normalidad

```
by(excope_18_p1$sumT, excope_18_p1$problema, shapiro.test)
```

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```
fligner.test(excope_18_p1$sumT, excope_18_p1$problema)
```

# Mantenemos el supuesto de homocedasticidad, pero como hay outliers y no se cumple el de normalidad, usamos la prueba robusta:

```

t1way(sumT ~ problema, excope_18_p1)

# Mantenemos H0. No hemos obtenido evidencia de que la solución al ítem del TRC
se asocie con diferencias en las puntuaciones a los cuatro ítems de la p1.

# Para finalizar con el análisis de influencia de las VI sobre las puntuaciones en el
experimento, analizamos la influencia conjunta de las variables "etiqueta" y "orden".

table(excope_18_p1$etiqueta, excope_18_p1$orden)

# Como se ha señalado, en la tabla vemos que no hay ningún sujeto que haya
respondido a un formato ordenado de acuerdo a desacuerdo en combinación con no
estar etiquetado. Pero vemos que hay 217 personas que han respondido en orden
"da" en una escala etiquetada.

aggregate(sumT ~ etiqueta + orden, summary, data = excope_18_p1)
aggregate(sumT ~ etiqueta + orden, sd, data = excope_18_p1)

eo1 <- excope_18_p1[excope_18_p1$orden == "da",]

# Análisis descriptivo:

describeBy(eo1$sumT, eo1$etiqueta)
ggplot(eo1, aes(factor(etiqueta), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas para orden = da")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores

rp.outlier(eo1[eo1$etiqueta == "etiquetado", "sumT"])
rp.outlier(eo1[eo1$etiqueta == "no_etiquetado", "sumT"])

# Comprobación del supuesto de normalidad

by(eo1$sumT, eo1$etiqueta, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(eo1$sumT, eo1$etiqueta)

# Mantenemos el supuesto de homocedasticidad, pero como parece haber outliers y
no se cumple el supuesto de normalidad, utilizaremos la prueba robusta:

yuen(sumT ~ etiqueta, eo1)

```

# Rechazamos H0. El resultado indica que hay diferencias en las puntuaciones a los cuatro ítems debidas a la presencia o no de etiqueta entre los que han recibido las categorías de respuesta en orden “da”.

# Figura 1a:

```
g1 <- ggplot(excope_18_p1, aes(factor(opciones), sumT)) +  
  geom_boxplot()  
g2 <- ggplot(excope_18_p1, aes(factor(parimpar), sumT)) +  
  geom_boxplot()  
g3 <- ggplot(excope_18_p1, aes(factor(etiqueta), sumT)) +  
  geom_boxplot()  
g4 <- ggplot(excope_18_p1, aes(factor(polaridad), sumT)) +  
  geom_boxplot()  
g5 <- ggplot(excope_18_p1, aes(factor(extremo), sumT)) +  
  geom_boxplot()  
g6 <- ggplot(excope_18_p1, aes(factor(orden), sumT)) +  
  geom_boxplot()  
g7 <- ggplot(excope_18_p1, aes(factor(muestra), sumT)) +  
  geom_boxplot()  
g8 <- ggplot(excope_18_p1, aes(factor(problema), sumT)) +  
  geom_boxplot()  
grid.arrange(g1, g2, g3, g4, g5, g6, g7, g8, ncol = 2)
```

## ANEXO IV – Análisis de la variable sumT en la pregunta 2

```
# Variable PUNTUACIONES TRANSFORMADAS:
```

```
# Fichero de datos: excope_18
```

```
# Cargamos los datos
```

```
load("J:/Privada/Ana/Máster Estadística con R/TFM/Código/excope_18.RData")
```

```
# HIPÓTESIS:
```

```
# H0: La media de las puntuaciones transformadas es igual para los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
# H1: La media de las puntuaciones transformadas difiere entre los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
#1) PREGUNTA 2
```

```
# Creamos el fichero con los datos para la pregunta 2:
```

```
excope_18_p2 <- excope_18[excope_18$pregunta == "p2",]
```

```
# Análisis descriptivo VD:
```

```
summary(excope_18_p2$sumT)
```

```
sd(excope_18_p2$sumT)
```

```
hist(excope_18_p2$sumT,  
      main = "Histograma P2- Suma P. Transformadas",  
      xlab = "Puntuaciones transformadas")
```

```
# VI: opciones
```

```
# Análisis descriptivo:
```

```
library(psych)
```

```
describeBy(excope_18_p2$sumT, excope_18_p2$opciones)
```

```
library(ggplot2)
```

```
ggplot(excope_18_p2, aes(factor(opciones), sumT)) +  
  geom_boxplot() +  
  labs(title = "Puntuaciones transformadas P2 - Opciones")
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
# El gráfico de cajas indica que hay dos. Comparando con la p1, parecen poco relevantes.
```

```
# Para identificar los valores:
```

```

library(rapportools)

rp.outlier(excope_18_p2[excope_18_p2$opciones == "4", "sumT"])
rp.outlier(excope_18_p2[excope_18_p2$opciones == "5", "sumT"])
rp.outlier(excope_18_p2[excope_18_p2$opciones == "6", "sumT"])

# La función rp.outlier no identifica valores extremos

# Comprobación del supuesto de normalidad

by(excope_18_p2$sumT, excope_18_p2$opciones, shapiro.test)

# Rechazamos la asunción del supuesto de normalidad para los tres grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando la prueba de Flinger-Killeen (no paramétrica)

fligner.test(excope_18_p2$sumT, excope_18_p2$opciones)

# Mantenemos el supuesto de homogeneidad de varianzas.

# Aunque se cumple el supuesto de homocedasticidad, no se cumple el de normalidad
y podría haber outliers, así que utilizaremos la prueba robusta.

library(WRS2)
t1way(sumT ~ opciones, excope_18_p2)

# Rechazamos H0. Hay evidencia de diferencias en las puntuaciones de los ítems
de la pregunta 2 relacionadas con el número de opciones de la escala de respuesta. El
tamaño del efecto es moderado.

# Comparaciones múltiples:

lincon(sumT ~ opciones, excope_18_p2)

# VI: parimpar

# Análisis descriptivo:

describeBy(excope_18_p2$sumT, excope_18_p2$parimpar)
ggplot(excope_18_p2, aes(factor(parimpar), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Parimpar")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que hay uno. Comprobamos para identificar los valores

```

```

rp.outlier(excope_18_p2[excope_18_p2$parimpar == "par", "sumT"])
rp.outlier(excope_18_p2[excope_18_p2$parimpar == "impar", "sumT"])

# Constatamos que no hay

# Comprobación del supuesto de normalidad

by(excope_18_p2$sumT, excope_18_p2$parimpar, shapiro.test)

# Rechazamos la asunción del supuesto de normalidad para los dos grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Flinger-Killeen:

fligner.test(excope_18_p2$sumT, excope_18_p2$parimpar)

# La evidencia obtenida nos lleva a asumir el supuesto de homocedasticidad, pero
como hay outliers y no se cumple el de normalidad, utilizaremos la prueba robusta

yuen(sumT ~ parimpar, excope_18_p2)

# Rechazamos la hipótesis nula, a diferencia de lo que ocurría con la pregunta 1. El
tamaño del efecto no es muy grande, pero se puede considerar relevante.

#VI: etiqueta

# Análisis descriptivo:

describeBy(excope_18_p2$sumT, excope_18_p2$etiqueta)
ggplot(excope_18_p2, aes(factor(etiqueta), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Etiqueta")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que hay uno. Comprobamos para tratar de identificarlo.

rp.outlier(excope_18_p2[excope_18_p2$etiqueta == "etiquetado", "sumT"])
rp.outlier(excope_18_p2[excope_18_p2$etiqueta == "no_etiquetado", "sumT"])

# La función no identifica el que aparece en el gráfico.

# Comprobación del supuesto de normalidad

by(excope_18_p2$sumT, excope_18_p2$etiqueta, shapiro.test)

```

# No se cumple el supuesto de normalidad, por lo que comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```
fligner.test(excope_18_p2$sumT, excope_18_p2$etiqueta)
```

# Mantenemos el supuesto de homocedasticidad, pero como hemos encontrado un posible outlier y no se cumple el de normalidad, utilizamos la prueba robusta

```
yuen(sumT ~ etiqueta, excope_18_p2)
```

# Rechazamos la H0 y, por tanto, hemos encontrado evidencia de que hay diferencias en la puntuación a los ítems de la p2 debidas a la presencia, o no, de etiquetas. La media es significativamente mayor cuando no se incluyen etiquetas. No obstante, hay que tener en cuenta que, tal y como se han diseñado las opciones en el experimento, hay coincidencia entre no tener etiqueta e ir en la dirección da en lugar de ad. El tamaño del efecto es moderado (0.34).

# VI: polaridad

# Análisis descriptivo:

```
describeBy(excope_18_p2$sumT, excope_18_p2$polaridad)
ggplot(excope_18_p2, aes(factor(polaridad), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Etiqueta")
```

# Comprobación de los supuestos:

# Identificación de outliers:

# El gráfico de cajas indica que no hay.

# Comprobación del supuesto de normalidad

```
by(excope_18_p2$sumT, excope_18_p2$polaridad, shapiro.test)
```

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```
fligner.test(excope_18_p2$sumT, excope_18_p2$polaridad)
```

# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad. Por tanto, utilizaremos la prueba robusta:

```
yuen(sumT ~ polaridad, excope_18_p2)
```

# No encontramos evidencia para rechazar H0.

# VI: extremo

```

# Análisis descriptivo:

describeBy(excope_18_p2$sumT, excope_18_p2$extremo)
ggplot(excope_18_p2, aes(factor(extremo), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Extremo")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que no hay.

# Comprobación del supuesto de normalidad

by(excope_18_p2$sumT, excope_18_p2$extremo, shapiro.test)

# No se cumple el supuesto de normalidad en los dos grupos, por lo que hay que
rechazarlo. Comprobaremos el supuesto de homogeneidad de varianzas utilizando el
test de Fligner-Killeen:

fligner.test(excope_18_p2$sumT, excope_18_p2$extremo)

# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad, por lo que
usamos la prueba robusta:

yuen(sumT ~ extremo, excope_18_p2)

# Mantenemos H0. No hay evidencia de que las puntuaciones en los 4 ítems de la p2
varíen de manera significativa debido a la etiqueta utilizada para definir el extremo de
la escala.

# VI: orden

# Análisis descriptivo:

describeBy(excope_18_p2$sumT, excope_18_p2$orden)
ggplot(excope_18_p2, aes(factor(orden), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Orden")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que hay uno. Comprobamos para identificar los valores

rp.outlier(excope_18_p2[excope_18_p2$orden == "ad", "sumT"])
rp.outlier(excope_18_p2[excope_18_p2$orden == "da", "sumT"])

# Comprobación del supuesto de normalidad

```

```

by(excope_18_p2$sumT, excope_18_p2$orden, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p2$sumT, excope_18_p2$orden)

# Mantenemos el supuesto de homocedasticidad, pero como parece haber outlier y no
se cumple el supuesto de normalidad, utilizaremos la prueba robusta:

yuen(sumT ~ orden, excope_18_p2)

# Rechazamos H0. Hay evidencia de que el orden en el que se describe la escala
influye en las puntuaciones medias. El tamaño del efecto es moderado a alto, mayor
que el encontrado para la pregunta 1. En todo caso, al tratarse de escalas inversas, lo
que el resultado nos está indicando es que personas que han respondido a los ítems
de la p2 tienden a mostrarse de acuerdo con las afirmaciones. Se ve muy bien en el
gráfico de cajas.

# De todos modos, visto el resultado obtenido con la variable etiqueta y la variable
orden, conviene analizar el efecto conjunto de ambas variables.

# VI: Muestra

# Análisis descriptivo:

describeBy(excope_18_p2$sumT, excope_18_p2$muestra)
ggplot(excope_18_p2, aes(factor(muestra), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Muestra")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que no hay

# Comprobación del supuesto de normalidad

by(excope_18_p2$sumT, excope_18_p2$muestra, shapiro.test)

# Como no se cumple el supuesto de normalidad en los dos grupos, comprobaremos
el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p2$sumT, excope_18_p2$muestra)

# Rechazamos el supuesto de homocedasticidad y el de normalidad, utilizaremos la
prueba robusta:

yuen(sumT ~ muestra, excope_18_p2)

```

# Mantenemos la H0. No hemos obtenido evidencia de que haya diferencias en las puntuaciones entre la muestra de la población general y la muestra del Ciemat. No debemos olvidar que se trata de un conjunto de voluntarios muy interesados por la ciencia. A diferencia de lo que ocurría con la p1, no se detectan casos de estilo extremo.

# VI: problema

# Análisis descriptivo:

```
describeBy(excope_18_p2$sumT, excope_18_p2$problema)
ggplot(excope_18_p2, aes(factor(problema), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas P2 - Problema")
```

# Comprobación de los supuestos:

# Identificación de outliers:

# El gráfico de cajas indica que no hay.

# Comprobación del supuesto de normalidad

```
by(excope_18_p2$sumT, excope_18_p2$problema, shapiro.test)
```

# Como no se cumple el supuesto de normalidad en los tres grupos, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```
fligner.test(excope_18_p2$sumT, excope_18_p2$problema)
```

# Rechazamos el supuesto de homocedasticidad y no se cumple el de normalidad, usamos la prueba robusta, que no es sensible al incumplimiento de homocedasticidad, algo que sí ocurre con la prueba paramétrica:

```
t1way(sumT ~ problema, excope_18_p2)
```

# Mantenemos H0. No hemos obtenido evidencia de que la solución al ítem del TRC se asocie con diferencias en las puntuaciones a los cuatro ítems de la p2.

# Para finalizar con el análisis de influencia de las VI sobre las puntuaciones en el experimento, analizamos la influencia conjunta de las variables "etiqueta" y "orden".

```
table(excope_18_p2$etiqueta, excope_18_p2$orden)
```

# Como se ha señalado, en la tabla vemos que no hay ningún sujeto que haya respondido a un formato ordenado de acuerdo a desacuerdo en combinación con no estar etiquetado. Pero vemos que hay 201 personas que han respondido en orden "da" en una escala etiquetada.

```

aggregate(sumT ~ etiqueta + orden, summary, data = excope_18_p2)
aggregate(sumT ~ etiqueta + orden, sd, data = excope_18_p2)

eo2 <- excope_18_p2[excope_18_p2$orden == "da",]

# Análisis descriptivo:

table(eo2$etiqueta)
describeBy(eo2$sumT, eo2$etiqueta)
ggplot(eo2, aes(factor(etiqueta), sumT)) +
  geom_boxplot() +
  labs(title = "Puntuaciones transformadas para orden = da")

# Comprobación de los supuestos:
# Identificación de outliers:
# El gráfico de cajas indica que hay uno. Comprobamos para identificar los valores

rp.outlier(eo2[eo2$etiqueta == "etiquetado", "sumT"])
rp.outlier(eo2[eo2$etiqueta == "no_etiquetado", "sumT"])

# Comprobación del supuesto de normalidad

by(eo2$sumT, eo2$etiqueta, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(eo2$sumT, eo2$etiqueta)

# Mantenemos el supuesto de homocedasticidad, pero como parece haber outliers y
no se cumple el supuesto de normalidad, utilizaremos la prueba robusta:

yuen(sumT ~ etiqueta, eo2)

# Mantenemos H0. El resultado indica que no hay diferencias en las puntuaciones a
los cuatro ítems de la p2 debidas a la presencia o no de etiqueta entre los que han
recibido las categorías de respuesta en orden "da". En este caso, entonces, el efecto
encontrado por las etiquetas se debe, en realidad, al efecto del orden de presentación
de las categorías de la escala.

# Figura 1b:

g1 <- ggplot(excope_18_p2, aes(factor(opciones), sumT)) +
  geom_boxplot()
g2 <- ggplot(excope_18_p2, aes(factor(parimpar), sumT)) +
  geom_boxplot()
g3 <- ggplot(excope_18_p2, aes(factor(etiqueta), sumT)) +
  geom_boxplot()

```

```
g4 <- ggplot(excope_18_p2, aes(factor(polaridad), sumT)) +  
  geom_boxplot()  
g5 <- ggplot(excope_18_p2, aes(factor(extremo), sumT)) +  
  geom_boxplot()  
g6 <- ggplot(excope_18_p2, aes(factor(orden), sumT)) +  
  geom_boxplot()  
g7 <- ggplot(excope_18_p2, aes(factor(muestra), sumT)) +  
  geom_boxplot()  
g8 <- ggplot(excope_18_p2, aes(factor(problema), sumT)) +  
  geom_boxplot()  
grid.arrange(g1, g2, g3, g4, g5, g6, g7, g8, ncol = 2)
```

## **ANEXO V – Análisis de la variable “variabilidad” en la pregunta 1.**

```
# Variable dependiente: VARIABILIDAD
```

```
# Fichero de datos: excope_18
```

```
# Cargamos los datos
```

```
load("J:/Privada/Ana/Máster Estadística con R/TFM/Código/excope_18.RData")
```

```
# HIPÓTESIS:
```

```
# H0: La media de variabilidad es igual para los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
# H1: La media de variabilidad difiere entre los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
#1) PREGUNTA 1
```

```
# Creamos el fichero con los datos para la pregunta 1:
```

```
excope_18_p1 <- excope_18[excope_18$pregunta == "p1",]
```

```
# Análisis descriptivo VD:
```

```
summary(excope_18_p1$variabilidad)
```

```
sd(excope_18_p1$variabilidad)
```

```
length(excope_18_p1$variabilidad)
```

```
library(ggplot2)
```

```
ggplot(excope_18_p1, aes(x = variabilidad)) +
```

```
  geom_bar() +
```

```
  labs(title = "Variabilidad P1")
```

```
# VI: opciones
```

```
# Análisis descriptivo:
```

```
library(psych)
```

```
describeBy(excope_18_p1$variabilidad, excope_18_p1$opciones)
```

# Variabilidad tiene muy poca variación. Como hemos visto al hacer el análisis descriptivo, primer, segundo y tercer cuartil tienen el mismo valor (0.75). Esto quiere decir que la mitad de la muestra ha utilizado tres opciones de respuesta para responder a los ítems de la p1. Además, es una variable discreta con cuatro valores, como resultado de que solo tenemos 4 ítems. Por tanto, el gráfico de cajas no es apropiado para representar esta variable. Utilizaremos el diagrama de barras

```
ggplot(excope_18_p1, aes(x = variabilidad, fill = opciones)) +  
  geom_bar(position = "dodge") +  
  labs(title = "Variabilidad P1 - opciones")
```

# Por otro lado, con independencia del cumplimiento o no de los supuestos, por ser una variable discreta deberíamos utilizar las pruebas no paramétricas. Si no se cumpliera el supuesto de homocedasticidad, a cuyo incumplimiento son muy sensibles estas pruebas, utilizaremos la prueba robusta de manera adicional para asegurar el resultado.

# Comprobación de los supuestos:

# Identificación de outliers:

```
library(rapportools)
```

```
rp.outlier(excope_18_p1[excope_18_p1$opciones == "4", "variabilidad"])  
rp.outlier(excope_18_p1[excope_18_p1$opciones == "5", "variabilidad"])  
rp.outlier(excope_18_p1[excope_18_p1$opciones == "6", "variabilidad"])
```

# La función rp.outlier no identifica valores extremos

# Comprobación del supuesto de normalidad

```
by(excope_18_p1$variabilidad, excope_18_p1$opciones, shapiro.test)
```

# Rechazamos la asunción del supuesto de normalidad para los tres grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando la prueba de Flinger-Killeen (no paramétrica)

```
fligner.test(excope_18_p1$variabilidad, excope_18_p1$opciones)
```

# Rechazamos el supuesto de homogeneidad de varianzas.

# Como no se cumple el supuesto de homocedasticidad, no se cumple el de normalidad y hay outliers, utilizaremos la prueba robusta.

```
library(WRS2)
```

```
t1way(variabilidad ~ opciones, excope_18_p1)
```

# Rechazamos H0. Hay evidencia de que la variabilidad entre las respuestas a los cuatro ítems de la p1 difiere en función del número de opciones de la escala de respuesta. El tamaño del efecto es débil (0.23).

# Realizamos las comparaciones múltiples

```
lincon(variabilidad ~ opciones, excope_18_p1)
```

# Hay diferencias significativas entre 5 y 6 opciones con respecto a 4, pero no entre 5 y 6.

# VI: parimpar

# Análisis descriptivo:

```
describeBy(excope_18_p1$variabilidad, excope_18_p1$parimpar)
ggplot(excope_18_p1, aes(x = variabilidad, fill = parimpar)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - parimpar")
```

# Comprobación de los supuestos:

# Identificación de outliers:

```
rp.outlier(excope_18_p1[excope_18_p1$parimpar == "par", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$parimpar == "impar", "variabilidad"])
```

# Constatamos que no hay. En todo caso, son personas con estilo de respuesta extremo. Hay que identificarlas

# Comprobación del supuesto de normalidad

```
by(excope_18_p1$variabilidad, excope_18_p1$parimpar, shapiro.test)
```

# Rechazamos la asunción del supuesto de normalidad para los dos grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Flinger-Killeen:

```
fligner.test(excope_18_p1$variabilidad, excope_18_p1$parimpar)
```

# La evidencia obtenida nos lleva a rechazar el supuesto de homocedasticidad, no se cumple el de normalidad, utilizaremos la prueba robusta.

```
yuen(variabilidad ~ parimpar, excope_18_p1)
```

# Mantenemos la hipótesis nula. No hay diferencias en la variabilidad en función de la polaridad de la escala de respuesta.

#VI: etiqueta

# Análisis descriptivo:

```
describeBy(excope_18_p1$variabilidad, excope_18_p1$etiqueta)
ggplot(excope_18_p1, aes(x = variabilidad, fill = etiqueta)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - etiqueta")
```

```

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p1[excope_18_p1$etiqueta == "etiquetado", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$etiqueta == "no_etiquetado", "variabilidad"])

# La función no los identifica.

# Comprobación del supuesto de normalidad

by(excope_18_p1$variabilidad, excope_18_p1$etiqueta, shapiro.test)

# No se cumple el supuesto de normalidad, por lo que comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p1$variabilidad, excope_18_p1$etiqueta)

# Mantenemos el supuesto de homocedasticidad, pero no se cumple el de normalidad,
utilizamos la prueba no paramétrica

wilcox.test(variabilidad ~ etiqueta, excope_18_p1)

# Mantenemos H0.

# VI: polaridad

# Análisis descriptivo:

describeBy(excope_18_p1$variabilidad, excope_18_p1$polaridad)
ggplot(excope_18_p1, aes(x = variabilidad, fill = polaridad)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - polaridad")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p1[excope_18_p1$polaridad == "unipolar", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$polaridad == "bipolar", "variabilidad"])

# En este caso la función no los identifica

# Comprobación del supuesto de normalidad

by(excope_18_p1$variabilidad, excope_18_p1$polaridad, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```

```

fligner.test(excope_18_p1$variabilidad, excope_18_p1$polaridad)

# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad, no hay
outliers, utilizaremos la prueba no paramétrica:

wilcox.test(variabilidad ~ polaridad, excope_18_p1)

# Mantenemos H0.

# VI: extremo

# Análisis descriptivo:

describeBy(excope_18_p1$variabilidad, excope_18_p1$extremo)
ggplot(excope_18_p1, aes(x = variabilidad, fill = extremo)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - extremo")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p1[excope_18_p1$extremo == "totalmente", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$extremo == "muy", "variabilidad"])

# La función no los identifica

# Comprobación del supuesto de normalidad

by(excope_18_p1$variabilidad, excope_18_p1$extremo, shapiro.test)

# No se cumple el supuesto de normalidad en los dos grupos, por lo que hay que
rechazarlo. Comprobaremos el supuesto de homogeneidad de varianzas utilizando el
test de Fligner-Killeen:

fligner.test(excope_18_p1$variabilidad, excope_18_p1$extremo)

# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad, no hay
outliers, utilizamos la prueba no paramétrica:

wilcox.test(variabilidad ~ extremo, excope_18_p1)

# Mantenemos H0. No hay evidencia de que la variabilidad entre las puntuaciones en
los 4 ítems de la p1 difiera de manera significativa debido a la etiqueta utilizada para
definir el extremo de la escala.

# VI: orden

```

```
# Análisis descriptivo:
```

```
describeBy(excope_18_p1$variabilidad, excope_18_p1$orden)
ggplot(excope_18_p1, aes(x = variabilidad, fill = orden)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - orden")
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
rp.outlier(excope_18_p1[excope_18_p1$orden == "ad", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$orden == "da", "variabilidad"])
```

```
# Comprobación del supuesto de normalidad
```

```
by(excope_18_p1$variabilidad, excope_18_p1$orden, shapiro.test)
```

```
# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:
```

```
fligner.test(excope_18_p1$variabilidad, excope_18_p1$orden)
```

```
# Mantenemos el supuesto de homocedasticidad, pero no se cumple el supuesto de
normalidad, no hay outliers, utilizaremos la prueba no paramétrica:
```

```
wilcox.test(variabilidad ~ orden, excope_18_p1)
```

```
# Rechazamos H0. Hay evidencia de que, al igual que ocurre con la variable "sumT",
el orden en el que se presentan las opciones de respuesta influye en la variabilidad de
las respuestas ofrecidas por los participantes.
```

```
# VI: Muestra
```

```
# Análisis descriptivo:
```

```
describeBy(excope_18_p1$variabilidad, excope_18_p1$muestra)
ggplot(excope_18_p1, aes(x = variabilidad, fill = muestra)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - muestra")
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
rp.outlier(excope_18_p1[excope_18_p1$muestra == "general", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$muestra == "general", "variabilidad"])
```

```
# Comprobación del supuesto de normalidad
```

```

by(excope_18_p1$variabilidad, excope_18_p1$muestra, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p1$variabilidad, excope_18_p1$muestra)

# Mantenemos el supuesto de homocedasticidad, pero como no se cumple el de
normalidad y no hay outliers, utilizamos la prueba no paramétrica:

wilcox.test(variabilidad ~ muestra, excope_18_p1)

# Mantenemos la H0. No hemos obtenido evidencia de que haya diferencias en las
puntuaciones entre la muestra de la población general y la muestra del Ciemat. No
debemos olvidar que se trata de un conjunto de voluntarios muy interesados por la
ciencia. Por otro lado, en la muestra general se detectan más casos de estilo extremo
de respuesta.

# VI: problema

# Análisis descriptivo:

describeBy(excope_18_p1$variabilidad, excope_18_p1$problema)
ggplot(excope_18_p1, aes(x = variabilidad, fill = problema)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P1 - problema")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p1[excope_18_p1$problema == "incorrecto", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$problema == "correcto", "variabilidad"])
rp.outlier(excope_18_p1[excope_18_p1$problema == "abandono", "variabilidad"])

# No hay

# Comprobación del supuesto de normalidad

by(excope_18_p1$variabilidad, excope_18_p1$problema, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p1$variabilidad, excope_18_p1$problema)

# Mantenemos el supuesto de homocedasticidad, pero no se cumple el de normalidad
y no hay outliers, usamos la prueba no paramétrica:

```

```
kruskal.test(variabilidad ~ problema, excope_18_p1)
```

# Mantenemos H0. No hemos obtenido evidencia de que la solución al ítem del TRC se asocie con diferencias en la variabilidad de las puntuaciones a los cuatro ítems de la p1.

# En el caso de la variabilidad en la pregunta 1, hemos encontrado efecto del número de opciones y el orden en el que se presentan.

#Figura 2a:

```
g1 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = opciones)) + geom_bar(position = "dodge")
g2 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = parimpar)) + geom_bar(position = "dodge")
g3 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = etiqueta)) + geom_bar(position = "dodge")
g4 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = polaridad)) + geom_bar(position = "dodge")
g5 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = extremo)) + geom_bar(position = "dodge")
g6 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = orden)) + geom_bar(position = "dodge")
g7 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = muestra)) + geom_bar(position = "dodge")
g8 <- ggplot(excope_18_p1, aes(x = variabilidad, fill = problema)) + geom_bar(position = "dodge")
grid.arrange(g1, g2, g3, g4, g5, g6, g7, g8, ncol = 2)
```

## ANEXO VI – Análisis de la variable “variabilidad” en la pregunta 2

```
# Variable dependiente: VARIABILIDAD
```

```
# Fichero de datos: excope_18
```

```
# Cargamos los datos
```

```
load("J:/Privada/Ana/Máster Estadística con R/TFM/Código/excope_18.RData")
```

```
# HIPÓTESIS:
```

```
# H0: La media de variabilidad es igual para los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
# H1: La media de variabilidad difiere entre los grupos creados a partir de las variables independientes (opciones, parimpar, etiqueta, polaridad, extremo, orden, muestra y problema)
```

```
#1) PREGUNTA 1
```

```
# Creamos el fichero con los datos para la pregunta 1:
```

```
excope_18_p2 <- excope_18[excope_18$pregunta == "p2",]
```

```
# Análisis descriptivo VD:
```

```
summary(excope_18_p2$variabilidad)
```

```
sd(excope_18_p2$variabilidad)
```

```
length(excope_18_p2$variabilidad)
```

```
# A diferencia de lo que ocurría en la pregunta 1, en este caso hay diferencia entre el primer y el segundo cuartil y el tercero.
```

```
library(ggplot2)
```

```
ggplot(excope_18_p2, aes(x = variabilidad)) +
```

```
  geom_bar() +
```

```
  labs(title = "Variabilidad P2")
```

```
# VI: opciones
```

```
# Análisis descriptivo:
```

```
library(psych)
```

```
describeBy(excope_18_p2$variabilidad, excope_18_p2$opciones)
```

```
ggplot(excope_18_p2, aes(x = variabilidad, fill = opciones)) +
```

```
  geom_bar(position = "dodge") +
```

```
  labs(title = "Variabilidad P2 - opciones")
```

```

# Comprobación de los supuestos:
# Identificación de outliers:

library(rapportools)

rp.outlier(excope_18_p2[excope_18_p2$opciones == "4", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$opciones == "5", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$opciones == "6", "variabilidad"])

# La función rp.outlier no identifica valores extremos

# Comprobación del supuesto de normalidad

by(excope_18_p2$variabilidad, excope_18_p2$opciones, shapiro.test)

# Rechazamos la asunción del supuesto de normalidad para los tres grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando la prueba de Flinger-Killeen (no paramétrica)

fligner.test(excope_18_p1$variabilidad, excope_18_p1$opciones)

# Rechazamos el supuesto de homogeneidad de varianzas.

# Como no se cumple el supuesto de homocedasticidad, no se cumple el de
normalidad y hay outliers, utilizaremos la prueba robusta.

library(WRS2)
t1way(variabilidad ~ opciones, excope_18_p1)

# Rechazamos H0. Hay evidencia de que la variabilidad entre las respuestas a los
cuatro ítems de la p1 difiere en función del número de opciones de la escala de
respuesta. El tamaño del efecto es débil (0.19).

# Realizamos las comparaciones múltiples

lincon(variabilidad ~ opciones, excope_18_p1)

#Encontramos diferencias entre 4 y 6 y entre 5 y 6 opciones. No encontramos
diferencias entre 4 y 5. El signo es negativo, eso indica que la variabilidad es mayor al
aumentar el número de opciones.

# VI: parimpar

# Análisis descriptivo:

describeBy(excope_18_p2$variabilidad, excope_18_p2$parimpar)

```

```

ggplot(excope_18_p2, aes(x = variabilidad, fill = parimpar)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - parimpar")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p2[excope_18_p2$parimpar == "par", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$parimpar == "impar", "variabilidad"])

# Constatamos que no hay. En todo caso, son personas con estilo de respuesta
extremo. Hay que identificarlas

# Comprobación del supuesto de normalidad

by(excope_18_p2$variabilidad, excope_18_p2$parimpar, shapiro.test)

# Rechazamos la asunción del supuesto de normalidad para los dos grupos

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Flinger-Killeen:

fligner.test(excope_18_p2$variabilidad, excope_18_p2$parimpar)

# La evidencia obtenida nos lleva a rechazar el supuesto de homocedasticidad, no se
cumple el de normalidad, utilizaremos la prueba robusta.

yuen(variabilidad ~ parimpar, excope_18_p2)

# Mantenemos la hipótesis nula. No hay diferencias en la variabilidad en función de la
polaridad de la escala de respuesta.

#VI: etiqueta

# Análisis descriptivo:

describeBy(excope_18_p2$variabilidad, excope_18_p2$etiqueta)
ggplot(excope_18_p2, aes(x = variabilidad, fill = etiqueta)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - etiqueta")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p2[excope_18_p2$etiqueta == "etiquetado", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$etiqueta == "no_etiquetado", "variabilidad"])

# La función no los identifica.

```

```

# Comprobación del supuesto de normalidad

by(excope_18_p2$variabilidad, excope_18_p2$etiqueta, shapiro.test)

# No se cumple el supuesto de normalidad, por lo que comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p2$variabilidad, excope_18_p2$etiqueta)

# Mantenemos el supuesto de homocedasticidad, pero no se cumple el de normalidad,
utilizamos la prueba no paramétrica

wilcox.test(variabilidad ~ etiqueta, excope_18_p2)

# Mantenemos H0.

# VI: polaridad

# Análisis descriptivo:

describeBy(excope_18_p2$variabilidad, excope_18_p2$polaridad)
ggplot(excope_18_p2, aes(x = variabilidad, fill = polaridad)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - polaridad")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p2[excope_18_p2$polaridad == "unipolar", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$polaridad == "bipolar", "variabilidad"])

# La función no los identifica

# Comprobación del supuesto de normalidad

by(excope_18_p2$variabilidad, excope_18_p2$polaridad, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p2$variabilidad, excope_18_p2$polaridad)

# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad, no hay
outliers, utilizaremos la prueba no paramétrica:

wilcox.test(variabilidad ~ polaridad, excope_18_p1)

```

```
# Mantenemos H0.
```

```
# VI: extremo
```

```
# Análisis descriptivo:
```

```
describeBy(excope_18_p2$variabilidad, excope_18_p2$extremo)
ggplot(excope_18_p2, aes(x = variabilidad, fill = extremo)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - extremo")
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
rp.outlier(excope_18_p2[excope_18_p1$extremo == "totalmente", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p1$extremo == "muy", "variabilidad"])
```

```
# La función no los identifica
```

```
# Comprobación del supuesto de normalidad
```

```
by(excope_18_p2$variabilidad, excope_18_p2$extremo, shapiro.test)
```

```
# No se cumple el supuesto de normalidad en los dos grupos, por lo que hay que rechazarlo. Comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:
```

```
fligner.test(excope_18_p2$variabilidad, excope_18_p2$extremo)
```

```
# Mantenemos el supuesto de homocedasticidad, pero no el de normalidad, no hay outliers, utilizamos la prueba no paramétrica:
```

```
wilcox.test(variabilidad ~ extremo, excope_18_p2)
```

```
# Mantenemos H0. No hay evidencia de que la variabilidad entre las puntuaciones en los 4 ítems de la p2 difiera de manera significativa debido a la etiqueta utilizada para definir el extremo de la escala.
```

```
# VI: orden
```

```
# Análisis descriptivo:
```

```
describeBy(excope_18_p2$variabilidad, excope_18_p2$orden)
ggplot(excope_18_p2, aes(x = variabilidad, fill = orden)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - orden")
```

```
# Comprobación de los supuestos:
```

```

# Identificación de outliers:

rp.outlier(excope_18_p2[excope_18_p2$orden == "ad", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$orden == "da", "variabilidad"])

# Comprobación del supuesto de normalidad

by(excope_18_p2$variabilidad, excope_18_p2$orden, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p2$variabilidad, excope_18_p2$orden)

# Mantenemos el supuesto de homocedasticidad, pero no se cumple el supuesto de
normalidad, no hay outliers, utilizaremos la prueba no paramétrica:

wilcox.test(variabilidad ~ orden, excope_18_p2)

# Rechazamos H0. Hay evidencia de que, al igual que ocurre con la variable "sumT",
el orden en el que se presentan las opciones de respuesta influye en la variabilidad de
las respuestas ofrecidas por los participantes.

# VI: Muestra

# Análisis descriptivo:

describeBy(excope_18_p2$variabilidad, excope_18_p2$muestra)
ggplot(excope_18_p2, aes(x = variabilidad, fill = muestra)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - muestra")

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18_p2[excope_18_p2$muestra == "general", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$muestra == "general", "variabilidad"])

# Comprobación del supuesto de normalidad

by(excope_18_p2$variabilidad, excope_18_p2$muestra, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18_p2$variabilidad, excope_18_p2$muestra)

```

# Mantenemos el supuesto de homocedasticidad, pero como no se cumple el de normalidad y no hay outliers, utilizamos la prueba no paramétrica:

```
wilcox.test(variabilidad ~ muestra, excope_18_p2)
```

# Mantenemos la H0. No hemos obtenido evidencia de que haya diferencias en las puntuaciones entre la muestra de la población general y la muestra del Ciemat. No debemos olvidar que se trata de un conjunto de voluntarios muy interesados por la ciencia. Por otro lado, en la muestra general se detectan más casos de estilo extremo de respuesta.

# VI: problema

# Análisis descriptivo:

```
describeBy(excope_18_p2$variabilidad, excope_18_p2$problema)
ggplot(excope_18_p2, aes(x = variabilidad, fill = problema)) +
  geom_bar(position = "dodge") +
  labs(title = "Variabilidad P2 - problema")
```

# Comprobación de los supuestos:

# Identificación de outliers:

# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores

```
rp.outlier(excope_18_p2[excope_18_p2$problema == "incorrecto", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$problema == "correcto", "variabilidad"])
rp.outlier(excope_18_p2[excope_18_p2$problema == "abandono", "variabilidad"])
```

# No hay

# Comprobación del supuesto de normalidad

```
by(excope_18_p2$variabilidad, excope_18_p2$problema, shapiro.test)
```

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de homogeneidad de varianzas utilizando el test de Fligner-Killeen:

```
fligner.test(excope_18_p2$variabilidad, excope_18_p2$problema)
```

# Mantenemos el supuesto de homocedasticidad, pero no se cumple el de normalidad y no hay outliers, usamos la prueba no paramétrica:

```
kruskal.test(variabilidad ~ problema, excope_18_p2)
```

# Rechazamos H0. La evidencia indica que la solución al ítem del TRC se asocia con diferencias en la variabilidad de las puntuaciones a los cuatro ítems de la p2.

# Realizamos las comparaciones múltiples:

```
library(pgirmess)
kruskalmc(variabilidad ~ problema, excope_18_p2)
```

#Hay diferencias entre incorrecto y abandono y entre correcto y abandono. No hay diferencias entre incorrecto y correcto.

# En el caso de la variabilidad en la pregunta 2, hemos encontrado efecto del número de opciones y el orden en el que se presentan las opciones, como en la p1. Además, hemos encontrado diferencias significativas relacionadas con la resolución del problema del TRC.

#Figura 2b:

```
g1 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = opciones)) + geom_bar(position = "dodge")
g2 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = parimpar)) + geom_bar(position = "dodge")
g3 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = etiqueta)) + geom_bar(position = "dodge")
g4 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = polaridad)) + geom_bar(position = "dodge")
g5 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = extremo)) + geom_bar(position = "dodge")
g6 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = orden)) + geom_bar(position = "dodge")
g7 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = muestra)) + geom_bar(position = "dodge")
g8 <- ggplot(excope_18_p2, aes(x = variabilidad, fill = problema)) + geom_bar(position = "dodge")
grid.arrange(g1, g2, g3, g4, g5, g6, g7, g8, ncol = 2)
```

## ANEXO VII - Análisis de las diferencias entre las dos preguntas

#Figura 3:

```
#  
library(gridExtra)  
g1 <- ggplot(excope_18, aes(x = sumT)) +  
  geom_histogram(fill = "white", color = "black", binwidth = 5) +  
  facet_wrap(~pregunta) +  
  labs(title = "Transformadas - Pregunta")  
g2 <- ggplot(excope_18, aes(x = variabilidad)) +  
  geom_bar(fill = "white", color = "black") +  
  facet_wrap(~pregunta) +  
  labs(title = "Variabilidad - Pregunta")  
grid.arrange(g1, g2, ncol = 2)
```

```
# Análisis de las diferencias en sumT entre las dos preguntas  
describeBy(excope_18$sumT, excope_18$pregunta)
```

```
# Comprobación de los supuestos:
```

```
# Identificación de outliers:
```

```
# El gráfico de cajas indica que los hay. Comprobamos para identificar los valores
```

```
rp.outlier(excope_18[excope_18$pregunta == "p1", "sumT"])
```

```
rp.outlier(excope_18[excope_18$pregunta == "p2", "sumT"])
```

```
# La función no los identifica. Además, se ve que están muy cerca de los bigotes.
```

```
# Comprobación del supuesto de normalidad
```

```
by(excope_18$sumT, excope_18$pregunta, shapiro.test)
```

```
# Rechazamos el supuesto de normalidad
```

```
#Comprobación del supuesto de homocedasticidad mediante Fligner-Killeen porque no  
se cumple el supuesto de normalidad
```

```
fligner.test(excope_18$sumT, excope_18$pregunta)
```

```
#Rechazamos el supuesto de homocedasticidad. Utilizamos la prueba robusta
```

```
yuen(variabilidad ~ pregunta, excope_18)
```

```
#Rechazamos H0. Hay diferencias en las puntuaciones a ambas preguntas. El tamaño  
del efecto es medio-alto (0.46).
```

```
# Comparación entre p1 y p2 en variabilidad:
```

```
describeBy(excope_18$variabilidad, excope_18$pregunta)
```

```

# Observamos que hay más variabilidad en la pregunta 1

# Comprobación de los supuestos:
# Identificación de outliers:

rp.outlier(excope_18[excope_18$pregunta == "p1", "variabilidad"])
rp.outlier(excope_18[excope_18$pregunta == "p2", "variabilidad"])

# No hay

# Comprobación del supuesto de normalidad

by(excope_18$variabilidad, excope_18$pregunta, shapiro.test)

# Como no se cumple el supuesto de normalidad, comprobaremos el supuesto de
homogeneidad de varianzas utilizando el test de Fligner-Killeen:

fligner.test(excope_18$variabilidad, excope_18$pregunta)

# Rechazamos el supuesto de homocedasticidad, por lo que utilizamos la prueba
robusta (no tenemos el mismo número de observaciones, por lo que no podemos
utilizar la prueba para muestras relacionadas, y tiene lógica porque los sujetos, de
manera mayoritaria, han respondido de manera independiente a ambas preguntas):

yuen(variabilidad ~ pregunta, excope_18)

#Hay diferencias significativas en la variabilidad entre ambas preguntas. El efecto es
medio-alto (0.43)

```