

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337826972>

# Investigación cuantitativa

Book · December 2018

---

CITATIONS

0

READS

2,249

1 author:



**Julián Cárdenas**

University of Valencia

37 PUBLICATIONS 406 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Business elites in Latin America [View project](#)



Ambiental psychology [View project](#)

# MANUAL DE CURSO

## Investigación cuantitativa

Julián Cárdenas

The DAAD logo consists of the letters 'DAAD' in a large, bold, blue, sans-serif font.

Con el apoyo financiero del



Ministerio Federal de  
Cooperación Económica  
y Desarrollo

Publicada por:

trAndeS – Programa de Posgrado en Desarrollo Sostenible y Desigualdades Sociales en la Región Andina

Lateinamerika-Institut, Freie Universität Berlin, Rüdeshheimer Str. 54-56, 14197 Berlin, Alemania

Esta publicación se ha realizado bajo los términos de la licencia Creative Commons 4.0 Attribution-NonCommercial-ShareAlike 4.0 International License ([CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)).

**trAndeS** *Serie de Material Docente* proporciona manuales de cursos que reúnen diferentes temas vinculados con el tema general del programa: la relación entre las desigualdades sociales y el desarrollo sostenible. Estos materiales buscan servir como recursos para la enseñanza y la formación.

El copyright de este manual pertenece a © Julián Cárdenas

Edición: Bettina Schorr / Paul Talcott / Frauke Berg

Todos los manuales de curso están disponibles gratuitamente en formato electrónico en nuestra página web [www.programa-trandes.net](http://www.programa-trandes.net).

Cárdenas, Julián 2018: “Investigación cuantitativa”, **trAndeS** Material Docente, No. 10, Berlín: **trAndeS** - Programa de Posgrado en Desarrollo Sostenible y Desigualdades Sociales en la Región Andina.

Las opiniones y puntos de vista expresados en este manual de curso son responsabilidad exclusiva del autor y no reflejan necesariamente las de **trAndeS**. **trAndeS** no es responsable por posibles errores ni por el uso de la información contenida en este manual de curso.

**Julián Cárdenas** es Doctor en Sociología. Profesor en técnicas de investigación social, análisis de redes, *networking* y sociología económica en diversas universidades de Alemania, España y Colombia. Investigador interesado en elites y redes. Puede conocer sus trabajos sobre redes en [www.networksprovidehappiness.com](http://www.networksprovidehappiness.com) y sobre técnicas de investigación en [www.networkianos.com](http://www.networkianos.com) Contacto: [julian.cardenas@onlinebschool.com](mailto:julian.cardenas@onlinebschool.com)

# Investigación cuantitativa

Julián Cárdenas

## Índice

Introducción.....	1
1. Tipos de métodos de investigación y principales técnicas de investigación .....	3
2. Cómo definir un problema de investigación.....	6
3. Cómo formular preguntas de investigación.....	10
4. Cómo identificar indicadores desde la pregunta de investigación pasando por las hipótesis. El proceso de operacionalización .....	15
5. Cómo y dónde recolectar datos de fuentes secundarias .....	19
6. Cómo hacer una encuesta y aplicarla .....	22
7. Cómo seleccionar y calcular muestras de una población .....	28
8. Cómo procesar y guardar los datos recolectados.....	32
9. Programas informáticos para análisis estadístico de datos cuantitativos .....	35
10. Análisis de datos: tipos de análisis estadísticos.....	36
11. Análisis univariado descriptivo .....	39
12. Cómo hacer y analizar tablas de contingencia.....	43
13. Correlaciones bivariadas.....	48
14. Qué es y cómo analizar ANOVA de un factor .....	54
15. Presentación de los resultados de investigación .....	60
Referencias .....	63

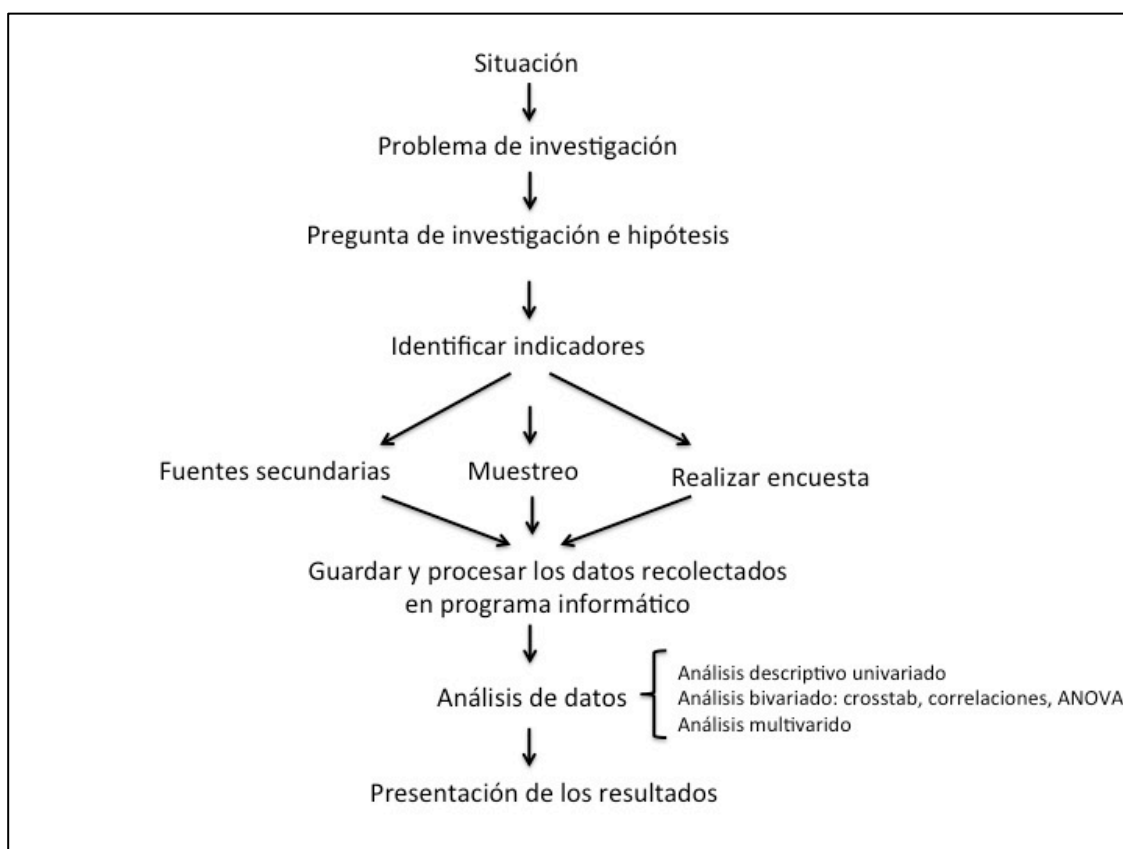
# Introducción

## El proceso de investigación

La investigación es un proceso destinado a resolver inquietudes a partir de la recolección y análisis de datos. La mejor manera de conocer e incluso controlar nuestra vida es investigando. La investigación produce conocimiento basado en datos que nos permite tomar decisiones más razonables, averiguar qué sucede más allá de nuestros ojos y tratar de predecir qué sucederá en el futuro. Hacer investigación nos hace más libres. Por todo ello merece la pena investigar.

La investigación requiere tiempo, recursos humanos y materiales, y sobre todo conocer el proceso de cómo se lleva a cabo una investigación. El siguiente cuadro representa de forma simplificada el proceso de investigación.

Figura 1. Proceso de investigación



Fuente: elaboración propia

¿Cuáles son las fases de una investigación? El origen de toda investigación está en el problema o asunto. El problema (también llamado asunto para no tener siempre una connotación negativa) es una situación concisa, relevante y contextualizada que interesa investigar. La pregunta de investigación es aquello que queremos averiguar sobre el problema y es la fase más relevante de una investigación ya que responderla será el objetivo o meta a conseguir. La pregunta de investigación se logra tras la consulta de bibliografía sobre el tema, la discusión con expertos en la materia, la observación crítica

de la realidad directa y la charla con actores implicados. Posteriormente, se identifican los indicadores que serán las formas en que se miden, cuantifican y observan los fenómenos de interés. Una vez que ya sabemos qué medir, se recolectan los datos, los cuales pueden ser obtenidos a partir de fuentes primarias (el investigador mismo genera los datos) o fuentes secundarias (los datos son generados por otros investigadores ajenos al proyecto). Una forma de generar los datos es mediante una encuesta por cuestionario. Esta técnica de investigación social es un conjunto de preguntas y categorías (u opciones) de respuestas aplicada a una muestra de actores. Los datos recolectados ya sean por encuesta o a través de fuentes secundarias son procesados y guardados en hojas de variables y datos. Para procesar, guardar y posteriormente analizar los datos se usan programas informáticos como Excel, SPSS, Stata, PSPP o R, entre otros. El análisis de los datos se desarrolla según el tipo de variables y el número de variables a analizar. Las técnicas estadísticas para analizar una variable son tablas de frecuencias, media y desviación estándar, y para analizar dos variables son tablas de contingencia, correlaciones y ANOVA de un factor, entre otras. El análisis de los datos tiene como meta responder las preguntas de investigación e identificar tendencias. Una vez analizados los datos, deben presentarse los resultados de forma escrita en un informe de investigación, y con ello culminamos el proceso de investigación. Parece difícil, ¿verdad? En realidad, es más sencillo de lo que parece. Y para mostrar que investigar es un proceso entretenido y abierto a todo tipo de personas, nace este manual.

El objetivo de este libro es que los lectores sean capaces de llevar a cabo una investigación siguiendo una metodología cuantitativa. Es un manual para conocer y practicar todas las fases del proceso de investigación, y para mostrar cómo analizar datos mediante técnicas estadísticas básicas. El presente libro se compone de 15 capítulos o sesiones. Cada sesión representa una fase de la investigación, y en ella se presentan las ideas principales, bibliografía sugerida y ejercicios prácticos. Este manual está dirigido a cualquier persona que tenga la necesidad o el deseo de investigar. La investigación debe estar al alcance de todos.

# 1. Tipos de métodos de investigación y principales técnicas de investigación

La investigación es un proceso de formular preguntas y responderlas mediante la recolección de datos. Estos datos pueden ser números, palabras o imágenes. Cuando la información son números (o bien la información recolectada es transformada en escalas numéricas) estamos ante una investigación con datos cuantitativos. Cuando la información son palabras o imágenes (no transformadas en escalas numéricas) estamos ante una investigación con datos cualitativos. Esta es la principal diferencia entre la investigación cuantitativa y cualitativa. En la investigación cuantitativa, los datos para responder las preguntas son números. En la investigación cualitativa, los datos recolectados son palabras o imágenes que no son sintetizadas en números.

El método es la parte del proceso de investigación centrada en cómo se responderán las preguntas de investigación. Detrás de los métodos hay una serie de paradigmas y teorías que señalan las fortalezas de cada uno. La decisión de optar por un método u otro depende de los objetivos de la investigación, los datos disponibles y obviamente de los recursos posibles para llevar a cabo un tipo de investigación u otra. Estos métodos de investigación también se conocen como enfoques de investigación.

Tabla 1. Diferencias entre método cuantitativo y cualitativo

Método cuantitativo	Método cualitativo
Basado en el positivismo lógico que pretende encontrar leyes que expliquen la realidad	Basado en la fenomenología y comprensión que pretende entender en profundidad el punto de vista de otros
Dirigido a datos medibles y cuantificables	Dirigido a las experiencias de los participantes
Suele usarse para objetivos de explicación	Suele usarse para objetivos de comprensión
Busca relaciones entre fenómenos	Busca la profundidad
Enfocadas en el resultado	Enfocadas en el proceso
Si se trabaja con muestras representativas, los resultados son generalizables a la población. Inferencia generalizable	Los resultados de la investigación no son generalizables a la población, aunque sí transferibles
Muchos casos	Pocos casos
Análisis estadístico	Análisis de contenido
Identificación de tendencias, comparación de grupos, relaciones entre variables	Identificación de categorías y descripción de temas
Datos numéricos	Datos en palabras o imágenes

Fuente: elaboración propia

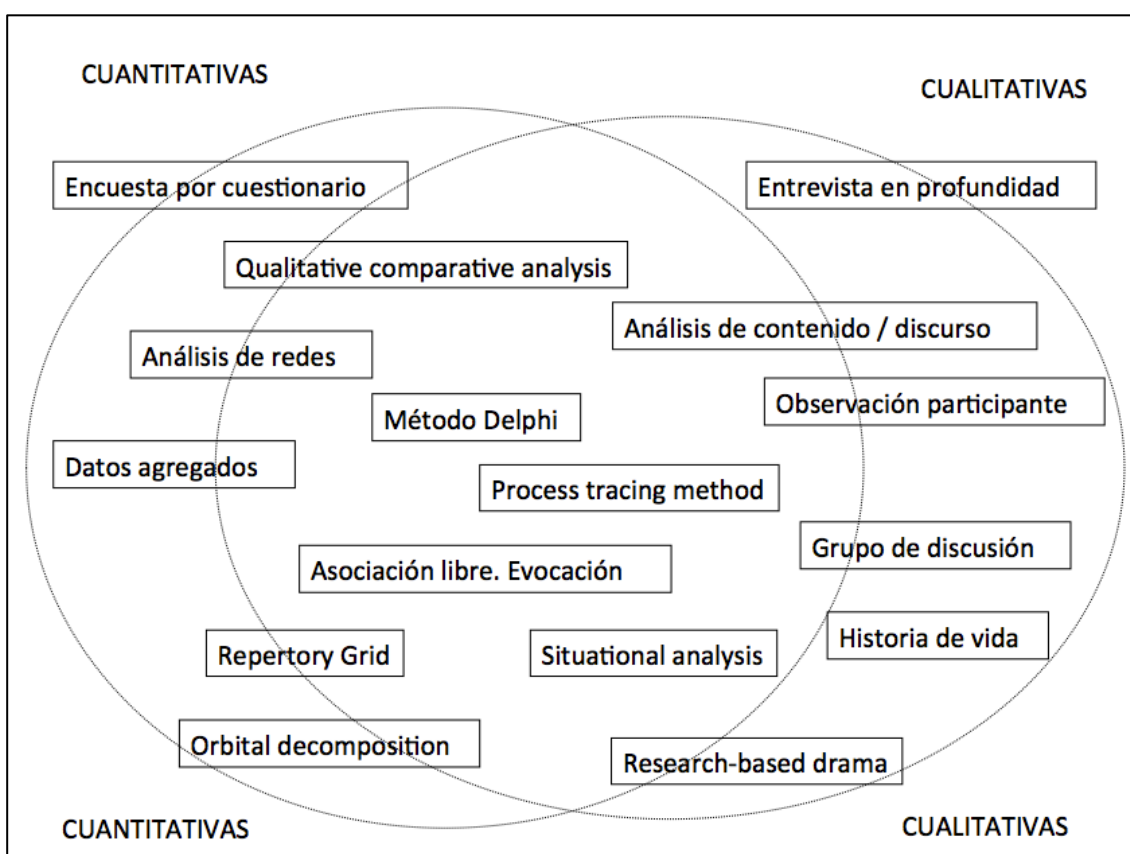
A pesar de estas distinciones entre el método cuantitativo y cualitativo, ambos enfoques se combinan e integran. Muchas investigaciones suelen ser una combinación de datos cuantitativos (números) y cualitativos (palabras o imágenes). La combinación de métodos de investigación cuantitativo y cualitativo se denomina metodología mixta. Por ejemplo, primero se recogen y analizan datos numéricos para conocer la extensión de un fenómeno y para seleccionar un segmento de la población. Posteriormente, se aplican entrevistas en profundidad a personas del segmento de población escogido y se contrastan las respuestas de los entrevistados con los datos numéricos de la población general. Por tanto, nunca



debe pensarse que los métodos cuantitativo y cualitativo son polos opuestos o contrincantes. La enseñanza de las metodologías por separado obedece más a procesos operativos que facilitan la comprensión y la organización. En la práctica, los métodos cuantitativo y cualitativo se combinan con mucha frecuencia.

La Figura 1 representa las principales técnicas de investigación social y su posición en el espectro de metodología cuantitativa y cualitativa. Se puede observar que hay un gran espacio de integración entre los métodos cuantitativo y cualitativo a la hora de aplicar las técnicas de investigación social, ya que para llevar a cabo una investigación se suelen usar elementos de ambas metodologías.

Figura 2. Técnicas de investigación social y tipos de métodos



Fuente: elaboración propia

### Bibliografía sugerida

Enfoque Cualitativo, Cuantitativo y Mixto

<https://www.youtube.com/watch?v=XpODjddaAqg>

### Ejercicios

Lea los siguientes artículos de investigación y responda las siguientes preguntas:

- ¿Qué tipo de método de investigación aplicaron los artículos: cuantitativo, cualitativo o mixto?
- Señale por qué cree que los investigadores utilizaron esa método o enfoque de investigación en cada uno de los artículos
- ¿Qué técnicas de investigación se aplicaron en cada uno de los artículos?

#### Artículo 1

Tobías-Martínez MA, Duarte-Freitas MC, Kemczinski A (2015) Un repositorio digital de contenido fílmico como recurso didáctico. *Comunicar* 44: 63-71

<http://www.revistacomunicar.com/indice/articulo.php?numero=44-2015-07>

#### Artículo 2

Chiappe-Laverde A, Hine N, Martínez-Silva JA (2015) Literatura y práctica: una revisión crítica acerca de los MOOC. *Comunicar* 44: 9-28.

<http://www.revistacomunicar.com/indice/articulo.php?numero=44-2015-01>

#### Artículo 3

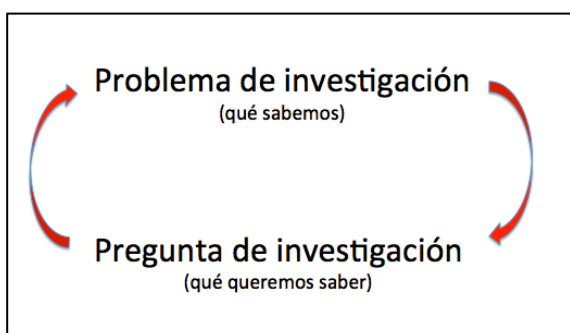
Öngün E, Demirag A. (2015). El uso de multimedias en las tareas académicas por los estudiantes. *Comunicar* 44: 121-129.

<http://www.revistacomunicar.com/indice/articulo.php?numero=44-2015-13>

## 2. Cómo definir un problema de investigación

El primer paso en una investigación es definir un problema de investigación, y el segundo paso es construir una pregunta de investigación a partir del problema de investigación. El problema de investigación es aquello que ya sabemos que ocurre y la pregunta de investigación es aquello que queremos averiguar o saber. El problema de investigación y la pregunta de investigación interactúan durante toda esta fase inicial hasta que logramos definir y delimitar el problema, y seguidamente logramos formular la pregunta de investigación (Figura 2). Hacerle preguntas al problema nos ayudará a definirlo y delimitarlo.

Figura 3. Problema y pregunta de investigación



Fuente: elaboración propia

### **El problema de investigación: cómo pasar de situaciones o ideas generales a un problema de investigación definido y delimitado**

Alrededor de nosotros suceden cientos de situaciones: tráfico de drogas, absentismo escolar, prostitución infantil, elecciones presidenciales, felicidad, asistencia a los estadios de fútbol. Algunas de estas situaciones nos interesa investigarlas para aumentar el conocimiento o para encontrarles una solución. Las situaciones son ideas muy generales que deben ser definidas o delimitadas para encontrar el problema específico a investigar. Hemos de transformar las situaciones o intereses generales en un problema de investigación definido y delimitado. Para ello seguiremos un protocolo de 2 grandes pasos.

#### **1. Describir la situación**

El primer paso es realizar una descripción amplia de la situación que observamos o conocemos. Hay que hacer una lista de todos los problemas que detectamos en esa situación mediante un análisis de los actores implicados, síntomas, visiones, prejuicios, factores causales, posibles consecuencias, aspectos del contexto, experiencias, teorías sugerentes. Son útiles en esta descripción imágenes, vídeos, observación directa de la situación, conversaciones con personas implicadas, con expertos, revisión de literatura sobre la situación y los problemas detectados. Por ejemplo, si la situación que nos llama la atención es que hay muchos mendigos en la calle, el siguiente paso para realizar la descripción es conocer qué políticas existen para solucionar el problema, qué se ha investigado sobre personas sin hogar (*homeless*), qué otros problemas hay asociados a la mendicidad como alcoholismo, soledad o redes informales, qué otros actores implicados hay alrededor de esta situación como instituciones de apoyo, familias o comerciantes. Además, cabe conocer qué piensan los mendigos, qué opinan los expertos, por qué la

gente los ayuda o no. Una vez que ya tenemos más conocimiento de la situación, hemos de realizar una lista de los posibles problemas a investigar. Por ejemplo, el descontento de los comerciantes con los mendigos, la alta solidaridad entre los mendigos de una misma zona, las políticas para mejorar la salud de los mendigos. Esta lista de problemas puede ser lo más extensa que se desee. En el siguiente paso, se irán definiendo y delimitando estos problemas detectados para finalmente seleccionar el problema en el que centraremos la investigación.

## 2. Definición y delimitación del problema de investigación

En este paso hemos de escoger uno de los problemas detectados en la situación y definir 7 aspectos de este problema. Estos aspectos son inquietudes que al responderlas nos ayudará a definir y delimitar el problema.

- Delimitación conceptual: hay de definir el problema de forma concisa. Para ello nos hemos de preguntar:
  - cómo se denomina a este problema
  - cómo se ha definido en otras investigaciones
  - qué actores están implicados en este problema

De esta forma se clarifica **qué investigar**.

Por ejemplo, si estamos interesados en las personas que viven en la calle, hemos de examinar que este problema se conoce como personas sin hogar o *homelessness*, y seguidamente hemos de definir este concepto en base a estudios previos.

- Relevancia del problema: hay que señalar por qué este problema es importante investigarlo. Para ello podemos examinar:
  - los riesgos del problema
  - los costes asociados, tanto económicos, humanos como políticos
  - si el problema ha aumentado en los últimos años o disminuido
  - si se ha expandido el problema en diversos lugares
  - si es un problema reconocido o desconocido.

De esta manera se consigue clarificar **por qué investigarlo**.

- Aplicaciones: hay que exponer qué posibles aplicaciones prácticas, teóricas y académicas, tiene investigar este problema. Para ello nos debemos preguntar:
  - qué impactos tendrá aumentar el conocimiento sobre este problema
  - a qué contribuiríamos sabiendo más sobre este problema o resolviéndolo

De esta manera se consigue aclarar **para qué investigarlo**.

- Delimitación física-geográfica. Hay que saber:
  - dónde ocurre este problema
  - dónde interesa investigarlo. El problema puede suceder en muchos lugares, pero hay que clarificar por qué me interesa un lugar concreto.

De esta manera se delimita **dónde investigarlo**.

- Delimitación temporal. Hay que exponer:
  - cuándo ocurre este problema
  - en qué momento, etapa u años interesa investigarlo. El problema puede suceder a lo largo del tiempo o ser eterno, pero hay que señalar por qué me interesa esa etapa o años específicos.

De esta manera se delimita **cuándo voy a investigarlo**.

- Antecedentes: este aspecto hace referencia a las investigaciones previas realizadas sobre el problema a investigar. Pueden ser investigaciones de otros países y otros años. Lo interesante es exponer:
  - qué se conoce de este problema, qué se sabe y qué no se sabe
  - cómo se lo ha investigado
  - qué debates existen en torno al tema

Así identificamos qué no se sabe todavía sobre este problema. De esta manera se define **cuál es el aporte de mi investigación y/o qué vacío que cubrirá**.

- Formulación interrogativa: hay que interrogar al problema sobre las causas, consecuencias, posibles soluciones, comparaciones con otros lugares u otros momentos del tiempo. Este aspecto es una forma de formular posibles preguntas de investigación, lo cual explicamos en el siguiente capítulo. De esta manera se delimitará **qué queremos averiguar o saber**

Se recomienda escribir todos estos aspectos a partir del problema seleccionado, y posteriormente evaluar si el problema está bien definido y delimitado, y si vale la pena investigarlo. Solo mediante la redacción del problema es posible la mejora y corrección. Un párrafo de máximo 10 líneas puede ser suficiente para saber si el problema está bien definido y delimitado.

### **Cómo saber si el problema está bien definido y delimitado**

Para que un problema de investigación esté bien definido y delimitado debe cumplir con las siguientes condiciones:

- Conciso: el problema debe estar claramente expuesto de forma escrita y debe poder ser entendido al leer su exposición
- Relevante: el problema debe tener una importancia para que valga la pena invertir recursos en adquirir conocimiento sobre ello. Esta relevancia puede ser muy relativa y subjetiva. Por ejemplo, puede parecer más importante estudiar la pobreza que los espacios de ocio en los parques. Sin embargo, la relevancia de una investigación viene determinada por la exposición de los costes, riesgos, ventajas, extensión y posibles contribuciones de la investigación. Es la sección donde se “vende” nuestra idea de investigación.
- Contextualizado: el problema debe estar definido en espacio, tiempo y antecedentes. La exposición de estos aspectos es lo que otorga una visión específica y al tiempo integral de lo que se quiere investigar.
- Específico: se recomienda que el problema de investigación sea algo concreto, ya que la investigación debe ser alcanzable, y tratar de abarcar demasiado es uno de los principales errores a la hora de definir y delimitar el problema de investigación.

Otra forma de evaluar si el problema está bien definido y delimitado es construir preguntas de investigación con el problema. En el siguiente capítulo se explicita cómo formular preguntas de investigación.

### **Otros recursos:**

Para describir las situaciones y poder detectar problemas de investigación son útiles algunos protocolos como el árbol de problemas, mapa de actores y análisis DAFO (debilidades, amenazas, fortalezas, oportunidades).

### **Bibliografía sugerida:**

Aprender LyX. Planteamiento del Problema, Definición - Partes del Planteamiento de un Problema. <https://www.youtube.com/watch?v=jnqONgntR6g>

### **Ejercicios:**

A partir de esta fotografía que representa una situación, defina y delimite un problema de investigación. Aplique las fases y protocolo sugerido. Después, compare los problemas de investigación detectados con la de otros colegas o compañeros a partir de la misma situación.

Enlace: <https://tinyurl.com/ydbd5s9o>

### **3. Cómo formular preguntas de investigación**

La pregunta de investigación es el primer paso para comenzar una investigación y es un reto por el que pasan todos los que inician una tesis, proyecto o informe de investigación. No podemos comenzar a redactar o preparar un proyecto de investigación si no tenemos pregunta de investigación identificada. No podemos escoger técnicas, teorías o datos si no tenemos pregunta de investigación. Si no tenemos identificada una pregunta de investigación divagaremos y perderemos mucho tiempo y energía. La pregunta de investigación es la meta que buscamos responder y será nuestra guía durante todo el proceso de investigación. Es mejor perder días, semanas o meses buscando una pregunta de investigación que iniciar la investigación sin tener una pregunta. Si la pregunta de investigación cambia una vez iniciada la investigación, deberemos reiniciar la investigación y volver a revisar el problema, las teorías y la revisión bibliográfica.

Seguir un protocolo de 3 pasos y aplicar seis estrategias es la manera más eficiente para formular preguntas de investigación y evitar estar perdido y sin rumbo durante la redacción del proyecto de investigación y el trabajo de campo. Pero antes de redactar la pregunta de investigación cabe tener en cuenta las condiciones que debe cumplir.

#### **Condiciones de una pregunta de investigación:**

La pregunta de investigación debe cumplir 3 condiciones indispensables:

- **Concisa:** lenguaje sencillo y claro. Cualquier persona, incluso sin formación en nuestro campo, debe entender la pregunta. Frases cortas y directas, nada de lenguaje pomposo y pretencioso.
- **Alcanzable:** la pregunta debe tener respuesta posible y la recolección de datos para responderla debe ser viable.
- **Relevante:** se debe defender la importancia de dedicar una investigación a responder dicha pregunta argumentando los beneficios e impactos que generarán los resultados de investigación: a nivel teórico, empírico y social.

Si no cumple alguna de estas condiciones, no es una pregunta digna que merezca una investigación seria y profesional.

#### **3 pasos para formular una pregunta de investigación**

Existe un protocolo a seguir para idear preguntas de investigación. Los 3 pasos son:

1°. Definir un problema de investigación: esta fase está detallada en el capítulo anterior. Por ejemplo, la felicidad no es un problema o asunto de investigación, es demasiado general, debe ser definido de forma más específica. Siguiendo con este ejemplo, hay lugares en América Latina con muchos problemas sociales y desigualdad pero su población se siente muy satisfecha y manifiesta ser feliz. Datos de la Encuesta Mundial de Valores o Happy Planet Index señalan que Panamá, Colombia, Venezuela, Ecuador, Costa Rica y Honduras son países con un nivel de felicidad muy alto en comparación con las regiones de Europa y Norteamérica.

2°. Delimitar el problema de investigación: se debe concretar y especificar claramente qué estudiaremos. La delimitación del problema pasa por especificar la acción que sucede y los actores implicados (pueden ser organismos, personas, organizaciones, ecosistemas, países). Además, se recomienda especificar el lugar y/o tiempo del problema. Una frase

debe resumir el problema o asunto a estudiar. Por ejemplo, el nivel de felicidad es alto en los países de América Latina entre los años 2000 y 2017.

3°. Aplicar seis estrategias al problema definido y delimitado para que surjan varias preguntas de investigación. Se trata de interrogar al problema de investigación sobre:

- **Descripción o exploración del problema:** si se sabe muy poco del problema o no existe información suficiente debe realizarse una investigación exploratoria o descriptiva. En este caso, las preguntas al problema pueden estar redactadas iniciando con “hasta qué punto” o bien “existe”. Ello nos permitirá averiguar si existe o no el problema, o qué magnitud tiene. Por ejemplo:

- ¿Hasta qué punto hay alto nivel de felicidad en los países de América Latina entre 2000 y 2017?

Las preguntas puramente descriptivas no suelen ser adecuadas para una investigación doctoral o un proyecto de investigación competitivo ya que solo detallan el problema, pero no lo explican, ni lo relacionan con otro fenómeno, ni lo comparan. Las siguientes cinco estrategias sí permiten formular preguntas para investigaciones explicativas, relacionales y comparativas.

- **Causas del problema:** preguntarse por una o varias causas del problema. Si son varias las posibles causas del problema, la pregunta puede ser redacta con “por qué” o bien “cuáles son las causas o razones que explican o influyen”, o bien “bajo qué condiciones”. Por ejemplo:

- ¿Por qué el nivel de felicidad es alto en los países de América Latina? O  
¿Qué factores explican el alto nivel de felicidad en los países de América Latina?

Si queremos averiguar si una causa concreta explica o influye el problema, entonces colocaremos al inicio la posible causa y posteriormente el problema de investigación.

- ¿El nivel de religiosidad influye sobre el nivel de felicidad en los países de América Latina?

Si lo que se pretende es evaluar una política o la introducción de un cambio, o bien ya se sabe que algo es la causa del problema, pero quiere conocerse el proceso que relaciona la causa con el problema, entonces se redactaría la pregunta usando el “cómo”. Por ejemplo:

- ¿Cómo ha afectado a los niveles de felicidad la política de aumento de los impuestos al consumo en los países de América Latina entre 2000 y 2017?
- ¿Cómo influye el capital social colectivo en el nivel de felicidad de la población en los países de América Latina?

- **Consecuencias del problema:** preguntarse por las consecuencias o impactos del problema en algún ámbito. Es recomendable colocar el problema al inicio de la pregunta y posteriormente añadir el fenómeno donde tiene impacto, efectos o consecuencias. Por ejemplo:

- ¿Qué efectos tiene el nivel de felicidad sobre el desarrollo del coeficiente intelectual de los jóvenes entre los 18 y 30 años en los países de América Latina?



- ¿Cómo afecta el alto nivel de felicidad al gasto empresarial en salud en los países de América Latina?
- ¿Impacta el nivel de felicidad sobre la productividad laboral en los países de América Latina?
- **Solución al problema o intervención sobre el problema:** pensar una solución al problema que investigamos o qué sucederá con el problema si realizamos una intervención. Preguntarse qué sucedería si aplicamos esa solución al problema delimitado o cómo afectaría una acción al problema.
  - ¿Si se aplicaran políticas de control de la natalidad se reducirían los índices de felicidad en los países de América Latina?
  - ¿El cambio de los horarios laborales en los escolares reduciría los índices de felicidad en los países de América Latina?

Este tipo de preguntas son más comunes en investigaciones que aplican método experimental, o realizan proyecciones, o bien son investigaciones de diseño o intervención social.

- **Comparación del problema entre lugares o perfiles:** preguntarse si el problema o asunto sucede en otro lugar, o si existen similitudes y diferencias entre grupos de personas. De esta manera formulamos preguntas de investigación para una investigación comparativa.
  - ¿Hay diferentes niveles de felicidad en los países de América Latina entre la población urbana y rural?
  - ¿Cuáles son las diferencias y similitudes de nivel de felicidad entre las clases trabajadoras de los países de América Latina?

A estas preguntas comparativas pueden añadirse preguntas sobre las causas o consecuencias. Por ejemplo:

- ¿Por qué en algunos países de América Latina hay mayor nivel de felicidad que en otros?
- **Comparación del problema en el tiempo:** preguntarse si el problema actual sucedía antes, o si el problema pasado sucede hoy en día, o bien si el problema que ocurría en una época sucedía en otra. Así formulamos preguntas para una investigación longitudinal o temporal.
  - ¿El alto nivel de felicidad en los países de América Latina ha sido constante en los últimos cien años?
  - ¿Cómo han cambiado (o evolucionado) los niveles de felicidad en los países de América Latina tras la crisis financiera de 2008?

Lo ideal es formular todas las preguntas de investigación posibles aplicando las seis estrategias siempre teniendo en cuenta las tres condiciones que deben cumplir las preguntas de investigación: concisa, alcanzable, relevante. Una vez escritas las diversas preguntas de investigación hay que someterlas a validación charlando con expertos y revisando si ya han sido ampliamente estudiadas. Si han sido excesivamente estudiadas nuestras preguntas de investigación, y ya se conocen las respuestas, y por tanto no suponen ninguna innovación, es mejor descartarlas y seguir buscando otras. De todas las preguntas formuladas podemos seleccionar una sola o bien varias, siempre y cuando estén

interconectadas. Es decir, podemos preguntarnos por las causas y consecuencias de un problema, y compararlo entre lugares y en el tiempo.

Una vez identificada la pregunta de investigación que queremos responder debemos añadirle los actores o sujetos de análisis (por ejemplo, personas, jóvenes entre 18-30 años, países), y el espacio físico y temporal que investigaré (por ejemplo, en los países de América Latina en 2017, en los países de América Latina en los últimos cien años, en los países de América Latina tras la crisis financiera de 2008). Por tanto, una pregunta de investigación definitiva usando la estrategia de consecuencias sería:

- ¿Cómo afecta el alto nivel de felicidad de los países de América Latina a la productividad laboral entre los años 2000 y 2017?

Otro ejemplo, combinando varias estrategias, en este caso la estrategia de causas y la estrategia de comparación entre lugares:

- ¿Por qué el nivel de felicidad de los jóvenes entre 18 y 30 años que residen en Colombia es más alto que el de los que residen en Perú entre los años 2000 y 2017?

Los principales errores en la formulación de la pregunta de investigación llegan al saltarse los pasos. No podemos pensar en soluciones o consecuencias si no tenemos bien delimitado el problema. Debemos seguir cada uno de los pasos en orden. La pregunta de investigación es el primer y más importante logro para tener éxito en la redacción del proyecto de investigación o tesis. Sin pregunta no hay proyecto de investigación ni investigación que arranque. La pregunta de investigación constituye redactada en infinitivo el objetivo de la investigación, por ejemplo:

- Pregunta de investigación: ¿Por qué el nivel de felicidad de los jóvenes entre 18 y 30 años que residen en Colombia es más alto que el de los que residen en Perú entre los años 2000 y 2017?
- Objetivo: Identificar las causas que explican por qué el nivel de felicidad de los jóvenes entre 18 y 30 años que residen en Colombia es más alto que el de los que residen en Perú entre los años 2000 y 2017

## **Hipótesis**

Las respuestas tentativas a la pregunta de investigación son las hipótesis. Son tentativas porque aún no están basadas en datos sino en conjeturas hechas a partir de la revisión de literatura y de la observación de la realidad. Las hipótesis son aquello que queremos probar si se cumple o se refuta tras la recolección y análisis de datos.

Por ejemplo, si nuestra pregunta de investigación es

- ¿Por qué el nivel de felicidad de los jóvenes entre 18 y 30 años que residen en Colombia es más alto que el de los que residen en Perú en 2017?

Posibles hipótesis basadas en el conocimiento del tema, tras la consulta de literatura y de datos estadísticos consultados serían:

- El mayor nivel de alfabetización de los jóvenes colombianos entre 18 y 30 años explica el mayor nivel de felicidad respecto a los jóvenes peruanos.
- La mayor participación de los jóvenes en organizaciones no gubernamentales explica que en Colombia la gente sea más feliz que en Perú.

La pregunta de investigación y las hipótesis nos ayudarán a identificar los indicadores que hemos de medir, lo cual se aborda en el siguiente capítulo

**Ejercicio:**

Aplique el protocolo para formular preguntas de investigación a partir del problema de investigación definido y delimitado en el capítulo anterior. Valore las preguntas de investigación de sus compañeros a partir del cumplimiento de las tres condiciones indispensables de las preguntas de investigación. De todas las preguntas de investigación, presente una para discutir con sus colegas o compañeros.

## **4. Cómo identificar indicadores desde la pregunta de investigación pasando por las hipótesis. El proceso de operacionalización**

El gran reto en una investigación es saber cómo vamos a medir las ideas o conceptos de nuestras preguntas e hipótesis. ¿Cómo se mide el rendimiento escolar? ¿Cómo se mide el desarrollo económico de un país? ¿Cómo se mide el desarrollo sostenible en educación de un país? ¿Y la desigualdad social entre los habitantes de un país? Los conceptos abstractos como rendimiento escolar, desarrollo económico, desarrollo sostenible o desigualdad social no pueden ser observados directamente, y por tanto deben ser observados y medidos mediante indicadores. Un indicador es una unidad medible y observable. El rendimiento escolar se mide mediante indicadores como promedio de calificaciones en un año, asistencia a clase y número de premios obtenidos. El desarrollo económico se mide mediante indicadores como Producto Interior Bruto (PIB) o PIB per cápita. El desarrollo sostenible en educación de un país se puede medir mediante indicadores como tasa de alfabetización en el país, número de estudiantes universitarios por habitante, número de profesores por personas escolarizadas. La desigualdad social de un país se mide por indicadores como el índice GINI y el ratio 90/10.<sup>1</sup> En la investigación de corte más cuantitativo, los indicadores propuestos suelen ir dirigidos a obtener números. En la investigación cualitativa también se usan indicadores, aunque van más dirigidos a la obtención de datos no-numéricos.

Cómo medimos los fenómenos o conceptos de interés es clave en una investigación porque es cómo observamos el mundo, qué miramos y cómo lo valoramos. Por tanto, el proceso de identificación de indicadores es una de las partes esenciales en la investigación. El paso de conceptos abstractos a indicadores medibles se denomina proceso de operacionalización.

El siguiente protocolo a seguir ayuda a pasar de la pregunta de investigación a los indicadores, el cual nos resolverá la incógnita de cómo vamos a medir y observar aquello que nos hemos preguntado. Los indicadores no son respuestas a las preguntas, sino los elementos que buscaremos en la realidad para observarla y medirla.

### **Protocolo para identificar indicadores en una investigación:**

#### **1. Formular hipótesis**

Las hipótesis son respuestas posibles a la pregunta de investigación. Por tanto, deben ser claras, relevantes y basadas en nuestro razonamiento a partir de la revisión de la literatura y la observación directa de la realidad. La función de redactar las hipótesis es para facilitar la identificación del concepto o conceptos principales que deseamos investigar. La hipótesis está compuesta por 1 o 2 conceptos. Cuando la hipótesis es descriptiva (descriptiva o comparativa) existe un solo concepto. Cuando la hipótesis es relacional (causal o asociativa) hay dos conceptos.

---

<sup>1</sup> El ratio 90/10 se calcula dividiendo el porcentaje de ingresos que controla el 10% más rico de la población y el porcentaje de ingresos que controla el 10% más pobre de la población.

- Por ejemplo, en la siguiente hipótesis, hay un solo concepto “calidad de vida”: *En Barcelona hay mayor calidad de vida que en Madrid.*
- En esta hipótesis, hay dos conceptos “calidad de vida” y “consumo de drogas”: *El aumento de la calidad de vida en las ciudades españolas supondrá un aumento del consumo de drogas en la población joven.*

## 2. Definir el/los conceptos

Se deben definir los conceptos usando los trabajos realizados hasta el momento y las teorías. Por ello, es necesario un proceso de lectura de las investigaciones que se han realizado hasta el momento. La definición siempre ha de ser explícita, no es conveniente que dejemos conceptos vagamente definidos. El investigador puede añadir sus impresiones para complementar las definiciones. El investigador es el responsable de decidir si toma las definiciones provenientes de las teorías, las complementa o las modifica. Una buena estrategia es definir el concepto de varias maneras y usar tantos indicadores como sea posible derivados de estas definiciones.

Varios autores han definido “calidad de vida” de la siguiente manera:

- Calidad de vida es definida como el bienestar personal derivado de la satisfacción o insatisfacción con áreas que son importantes para él o ella (Ferrans, 1990).
- Calidad de vida es la evaluación multidimensional, de acuerdo a criterios intrapersonales y socio-normativos, del sistema personal y ambiental de un individuo (Lawton, 2001).
- Calidad de vida es equivalente a la suma de los puntajes de las condiciones de vida objetivamente medibles en una persona, tales como salud física, condiciones de vida, relaciones sociales, actividades funcionales u ocupación (Urzúa M and Caqueo-Urizar, 2012).

Cuál de estas definiciones se usa en la investigación es una decisión del investigador, y está condicionada por el objetivo de su investigación y por aquello que se pueda medir. Es decir, debe haber concordancia en cómo se define el concepto y cómo se medirá posteriormente.

## 3. Romper el concepto en dimensiones

Hay conceptos muy generales y abstractos como “calidad de vida” que requieren ser desmembrados o divididos en dimensiones, las cuales son las diversas partes que componen un concepto. El concepto de “calidad de vida” se puede fragmentar o dividir en estas dimensiones:

- “calidad de vida en vivienda”
- “calidad de vida en salud”
- “calidad de vida en educación”
- “calidad de vida en medioambiente”
- “calidad de vida en economía”
- “calidad de vida en ocio”

El número de dimensiones establecidas por cada concepto depende del investigador, de las teorías que use y de su definición previa. Cuantas más dimensiones se establezcan por concepto, más completa será la investigación ya que estará abarcando más partes.

El proceso de romper el concepto en dimensiones se realiza teniendo en cuenta:

- trabajos anteriores y teorías
- teniendo en cuenta nuestra pregunta e hipótesis de investigación
- las dimensiones establecidas deben ser independientes entre sí

#### **4. Identificar indicadores para cada dimensión del concepto**

Cada dimensión debe ser medida por indicadores. El reto es encontrar cómo medimos cada una de las dimensiones. Para ello buscamos indicadores que nos sirvan de pistas para observar la realidad. Por ejemplo, ¿cómo medimos “calidad de vida en vivienda”? Posibles indicadores serían:

- metros cuadrados por número de habitante en la vivienda
- número de electrodomésticos en la vivienda
- servicio sanitario en la vivienda
- acceso a servicios públicos de electricidad en la vivienda
- acceso a servicios públicos de agua potable en la vivienda
- acceso a servicios públicos de gas en la vivienda

El mismo proceso debemos realizarlo para cada una de las dimensiones del concepto de calidad de vida. El conjunto de indicadores identificados serán las formas en que mediremos y observaremos “calidad de vida”. Los indicadores serán utilizados para comparar el nivel de calidad de vida y comprobar o refutar la hipótesis.

Es importante a la hora de identificar indicadores:

- la población de estudio: no es lo mismo medir calidad de vida investigando ciudades que empresas. Los indicadores pueden ser similares, pero se adaptan las medidas. Por ejemplo, la calidad de vida en ocio se puede medir en ciudades mediante el indicador “número de cines y teatros por número de habitantes en la ciudad”. En cambio, la calidad de vida en ocio en una organización se puede medir en empresas mediante el indicador “disponibilidad de un espacio para actividades recreativas como juegos, lectura o charlas dentro de la empresa”.
- los indicadores se materializarán en preguntas de cuestionario, entrevistas o en nuestra guía en observaciones o etnografías. Por ejemplo, el indicador “disponibilidad de un espacio para actividades recreativas como juegos, lectura o charlas dentro de la empresa” se preguntaría en un cuestionario: *En su lugar de trabajo actual, ¿hay algún espacio reservado para actividades recreativas como juegos, lectura o charlas? Sí – No – NS/NC.*
- Los indicadores son características de la realidad y deben ser claros, específicos, oportunos, referidos al concepto, comparables, disponibles, observables, medibles y cuantificables.

En la redacción de proyectos de investigación o en el mismo informe final o artículo puede ser incluido un cuadro que muestre este proceso de operacionalización. Por ejemplo:

Tabla 2. Definición operativa de calidad de vida

DIMENSIONES	INDICADORES
Bienestar físico	Salud general, atención sanitaria, memoria, dolor, sueño, visión, audición, movilidad
Bienestar material	Empleo, ingresos, propiedades, ahorros, condiciones de la vivienda, condiciones del lugar de trabajo, condiciones del servicio, jubilación
Relaciones interpersonales	Tiene amigos claramente identificados, contactos sociales positivos o gratificantes, relación de pareja, relaciones familiares, otras relaciones sociales
Desarrollo personal	Habilidades funcionales, habilidades académicas
Bienestar emocional	Ausencia de estrés y sentimientos negativos, autoconcepto, satisfacción

Fuente: elaboración propia.

El éxito en la búsqueda de indicadores reside en leer, leer y volver a leer literatura relacionada con el concepto. Cabe revisar cómo otras investigaciones han medido el concepto de interés. El gran número de investigaciones existentes en el mundo es nuestra mejor herramienta para la búsqueda de indicadores.

El proceso de operacionalización es posiblemente uno de los pasos más complicados de la investigación y a la vez el más relevante ya que todo depende de cómo medimos nuestras ideas. Todo se puede medir, y al medirlo haremos que exista, podremos describirlo, analizarlo, criticarlo y solucionarlo.

Una vez identificados los indicadores, el siguiente paso será o bien diseñar un instrumento (por ejemplo, una encuesta) para recoger los datos que hagan referencia a los indicadores, o bien recurrir a fuentes secundarias (por ejemplo, bases de datos online) para recoger los datos que hagan referencia a los indicadores identificados. Esto lo veremos en los siguientes capítulos.

### **Bibliografía sugerida**

Cea d'Ancona, M.A. (1999) Metodología Cuantitativa: Estrategias y Técnicas de Investigación Social. Capítulo 4. Madrid: Síntesis

<http://metodos-comunicacion.sociales.uba.ar/files/2014/04/Cea-D-Ancona-Cap-3-y-4.pdf>

Reguant, M., y Martínez-Olmo, F. (2014). Operacionalización de conceptos / variables. Barcelona: Dipòsit Digital de la UB

<http://diposit.ub.edu/dspace/bitstream/2445/57883/1/Indicadores-Repositorio.pdf>

### **Ejercicio**

Identifique indicadores a partir de la pregunta de investigación que realizó en el capítulo anterior. Cree una tabla como la Tabla 2 señalando las dimensiones y los indicadores.

## 5. Cómo y dónde recolectar datos de fuentes secundarias

En el proceso de investigación se pueden distinguir dos tipos de fuentes de datos: primarias y secundarias.

- Las fuentes primarias o datos primarios son aquellos recolectados o producidos por el propio investigador o grupo que realiza la investigación. Son datos obtenidos a primera mano. Cuando realizamos una entrevista o una encuesta y nosotros recolectamos la información estamos usando fuentes primarias.
- Las fuentes secundarias o datos secundarios son informaciones recolectadas o producidas por personas o instituciones ajenas a las que realizan la investigación. Cuando usamos datos estadísticos del Banco Mundial, cuando consultamos documentos escritos de un archivo o cuando consultamos información bibliográfica estamos usando fuentes secundarias.

En el presente capítulo se explicita cómo recolectar datos de fuentes secundarias.

### ¿Qué información o datos debemos recolectar?

Los datos que buscamos y recolectamos son referidos a los indicadores. Hemos de recolectar información o datos sobre los indicadores que hemos identificado en la fase anterior. Si recolectamos datos por el simple hecho de recolectar, puede llevarnos a un trabajo eterno y sin rumbo, por ello es conveniente tener una guía de lo que hemos de buscar. Esta guía son los indicadores identificados en la fase anterior. Si durante la recolección de datos nos percatamos de otros datos que pueden ser útiles, obviamente podemos recolectarlos para completar los indicadores, pero siempre es conveniente buscar datos en base a los indicadores que miden nuestros conceptos o ideas principales.

### ¿Qué datos de fuentes secundarias son más confiables?

En la era de las noticias falsas y la posverdad, no es fácil confiar en los datos de fuentes secundarias. Sin embargo, se utilizan tres criterios importantes para identificar qué datos de una fuente secundaria pueden ser confiables:

- a) Cuantas más veces una fuente secundaria haya sido usada en artículos publicados en revistas científicas, más confiable es la fuente y los datos.
- b) Datos publicados en revistas científicas de alto impacto están más legitimados ya que varios expertos revisan estas investigaciones.
- c) Cuanto más estatus y reputación tenga una institución, más legitimados y confiables son sus datos. En la siguiente sección se presentan fuentes secundarias de instituciones con alto prestigio y reputación.

### ¿Dónde buscar datos secundarios sobre los indicadores?

Existen cientos o miles de organizaciones que recolectan periódicamente datos sobre diversas temáticas. Internet es una fuente inmensamente rica, heterogénea y económica para obtener información. Por ello cabe conocer cuáles son las principales fuentes secundarias en nuestra temática o campo de estudio. El siguiente listado señala varias opciones y tipos de fuentes secundarias donde obtener datos empíricos.

Instituciones internacionales – son organismos que ponen a disposición datos donde la unidad de análisis son los países o los continentes. Son muy útiles para llevar a cabo investigaciones que pretenden comparar países o estudiar la evolución en el tiempo de un



fenómeno. Algunas de las principales instituciones internacionales con bases de datos accesibles son:

- Banco Mundial – posiblemente la mayor fuente de datos por países sobre indicadores políticos, económicos, sociales, de salud o medioambiente. Acceso libre y posibilidad de obtener datos históricos. <http://datos.bancomundial.org/>
- CIA Factbook – datos compilados por la Agencia Central de Inteligencia de Estados Unidos. Ofrece guías de cada país donde se presentan tanto datos cuantitativos como cualitativos sobre diversos temas. Acceso libre. <https://www.cia.gov/library/publications/the-world-factbook/>
- CEPAL – Comisión Económica regional de Naciones Unidas para América Latina. Contiene datos de diversos temas, sobre todo económicos y comerciales, y contiene datos basados en los censos y encuestas realizadas a nivel nacional y local. También realizan informes donde se sintetizan y analizan los datos compilados. Acceso libre. <http://www.cepal.org/es/datos-y-estadisticas>
- UNESCO Institute for Statistics (UIS) – fuente de datos que permite la comparación entre países y estudios longitudinales sobre educación, ciencia y tecnología, cultura y comunicación. Acceso libre. <http://www.uis.unesco.org/Pages/default.aspx>
- Web of Science de Thomson Reuters – fuente de datos sobre publicaciones de artículos y libros científicos. Algunas bibliotecas universitarias tienen acceso a esta fuente bibliométrica ya que es de pago y requiere suscripción. <http://isiknowledge.com/wos> Una fuente de datos sobre revistas científicas de acceso libre es Scimago: <http://www.scimagojr.com/>

Encuestas a nivel internacional – algunos grupos de investigación recolectan datos mediante una misma encuesta en varios países. Esto permite realizar investigaciones comparadas entre países y de forma longitudinal, así como analizar las opiniones de la población de un país concreto.

- Encuesta Mundial de Valores – contiene opiniones sobre valores, creencias y comportamientos sociales. En su página web pueden realizarse análisis en línea y mapas al seleccionar los temas de interés. También pueden descargarse libremente los datos en Excel, SPSS, R, Stata, SAS y ASCII. <http://www.worldvaluessurvey.org/WVSContents.jsp>
- Latinobarómetro – estudio de opinión pública que se realiza en 18 países de América Latina. Pueden realizarse análisis online y también descargar los datos completos para su posterior análisis. <http://www.latinobarometro.org/latOnline.jsp>
- Latin American Public Opinion Project (LAPOP) – encuestas realizadas en diversos países de América Latina sobre la opinión, actitudes y comportamientos de la población en diversos asuntos políticos, sociales y económicos <http://www.vanderbilt.edu/lapop/>
- European Social Survey – estudio de las opiniones sobre diversos asuntos sociales realizado en 36 países de Europa. Tiene una aplicación excelente para realizar análisis estadísticos sin necesidad de instalar ningún software <http://www.europeansocialsurvey.org/data/>

Instituciones y encuestas nacionales – son organismos, normalmente públicos donde se recopila información sobre la población de un país. Son útiles cuando se centra la investigación en un país, o en las regiones o localidades del país.

- Instituto de Estadística – suelen incluir información obtenida por el censo y encuestas de diversos temas. Para Perú: <https://www.inei.gob.pe/bases-de-datos/>
- Superintendencia o comisión del mercado – incluye datos sobre grandes empresas. Para Perú: <http://www.smv.gob.pe/>

Estos son solo algunos ejemplos de fuentes secundarias, existen miles más. Una opción para localizar fuentes secundarias de datos es revisar literatura e investigaciones previas y comprobar de dónde extrajeron la información para medir los indicadores de interés.

En una investigación se pueden y se suelen usar varias fuentes secundarias. Los datos obtenidos de ellas se combinan para poder contestar la pregunta de investigación y medir los conceptos mediante los indicadores. Por ejemplo, para medir la desigualdad, se usa el índice GINI, el cual se obtiene del Banco Mundial. Este dato señala que la desigualdad de ingreso en 2014 es 44.1 en Perú y 53.5 en Colombia, por tanto, hay más desigualdad en Colombia que en Perú. Para recoger información sobre las opiniones de las personas respecto de la desigualdad puede consultarse el Latinobarómetro, específicamente la pregunta referida a “¿Cuál es el principal problema del país?” en 2015. Este dato me señala que solo el 0.2% de los encuestados en Perú y el 2% de los encuestados en Colombia señalan que la inequidad es el principal problema del país.

Toda fuente secundaria usada en la investigación debe ser citada en la bibliografía.

### **¿Qué ventajas y desventajas tiene usar fuentes secundarias en lugar de fuentes primarias?**

Las principales ventajas son:

- acceso a más datos en menos tiempo y gastando menos recursos
- permite comparaciones en el tiempo (longitudinales) y entre varios países

Las principales desventajas son:

- el investigador pierde el control de la recolección de información
- los datos pueden no hacer referencia explícita al indicador que se desea medir
- algunas fuentes de información secundarias son poco fiables y válidas

De todas estas razones la que más influye es el coste o recursos disponible para llevar a cabo la investigación, por lo que si los fondos económicos son limitados se suele optar por datos secundarios, especialmente si quieren hacerse investigaciones a nivel macro.

En el siguiente capítulo veremos cómo diseñar por nosotros mismos un instrumento de recolección de información, específicamente una encuesta.

### **Ejercicios**

A partir de los indicadores identificados en el capítulo anterior, recolecte datos sobre ellos en base a información proveniente de fuentes secundarias. Pueden ser bases de datos oficiales, datos de internet, libros, archivos o cualquier tipo de fuente. Se recomienda consultar las fuentes secundarias presentadas en este capítulo.

## 6. Cómo hacer una encuesta y aplicarla

La encuesta es un instrumento para recolectar datos. Aplicamos una encuesta cuando queremos recoger datos empíricos que no están disponibles. Se conoce por encuesta a la técnica de investigación que, mediante un instrumento de preguntas y categorías de respuesta, y una muestra de actores, permite obtener datos cuantitativos. Se conoce como cuestionario, el conjunto de preguntas y categorías de respuesta.

Se puede usar un cuestionario ya existente (con leves adaptaciones) o crear uno nuevo. Lo más recomendable es usar cuestionarios existentes y adaptarlos mínimamente al lugar, tiempo y población a la que se dirige, ya que el coste en tiempo y recursos es mayor cuando lo creamos desde cero.

Si optamos por diseñar un cuestionario, estos son los principales pasos a seguir.

### De los indicadores a la redacción de un cuestionario

#### 1. Los indicadores

El primer paso para la construcción de un cuestionario es tener identificados los indicadores. Si queremos medir “calidad de vida” y algunos de sus indicadores son “acceso a gas”, “ingresos en el hogar” y “acceso a guarderías”, las preguntas del cuestionario van a ir dirigidas a obtener información sobre estos indicadores. Sin indicadores no podemos comenzar un a redactar un cuestionario.

#### 2. Redacción de preguntas

Para cada indicador se redacta al menos una pregunta.

- ¿Tiene acceso a gas la vivienda en la que habita?

También puede incluirse el indicador dentro de una pregunta con varias opciones.

- De los siguientes servicios, dígame por favor, ¿a cuáles tiene acceso en su vivienda?
  1. Luz eléctrica
  2. Agua potable
  3. Gas
  4. Energía solar
  5. Recogida de basuras

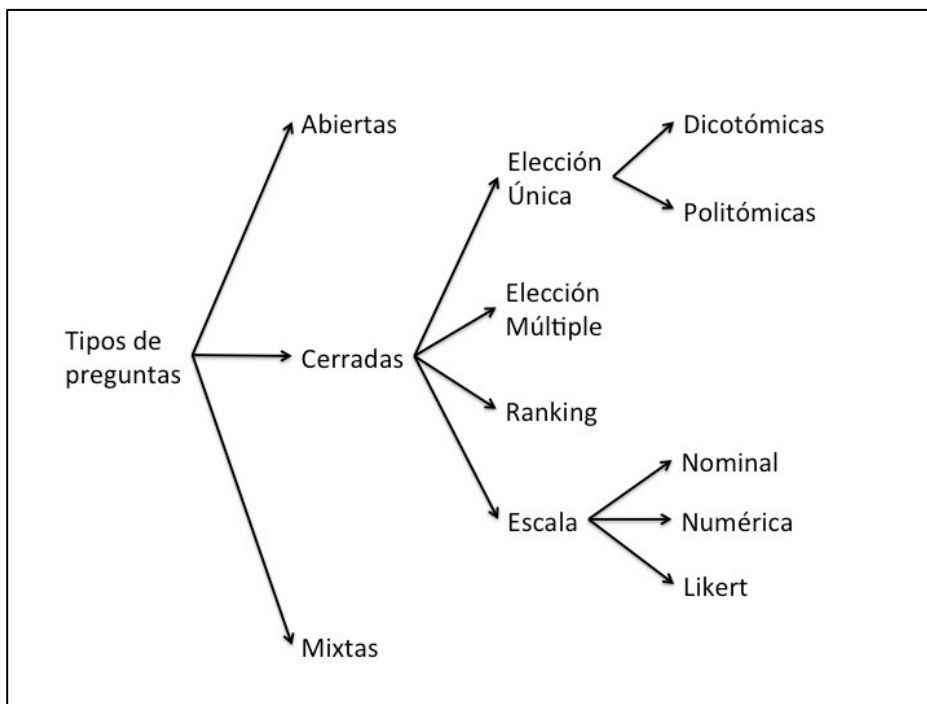
La mejor manera de aprender a redactar preguntas de cuestionario es revisar otros cuestionarios que hayan sido aplicados y tenga alta legitimidad, y así formular preguntas de forma similar. Se recomienda revisar la Encuesta Mundial de Valores, Latinobarómetro y el amplio catálogo de encuestas del Centro de Investigaciones Sociológicas de España ( [www.cis.es](http://www.cis.es) > Estudios). La redacción de las preguntas es uno de los procesos claves para que un cuestionario pueda validarse y aplicarse de forma exitosa.

Los **errores más habituales** al redactar preguntas de cuestionario y que debemos evitar son:

- Preguntas que combinen dos o más temas en el enunciado. Un ejemplo de pregunta mal redactada sería: ¿Qué grado de satisfacción tiene de la política educativa y económica del gobierno?
- Preguntas que requieren alto esfuerzo de memoria del encuestado. Ej. ¿Cuántas veces ha donado una limosna en los últimos dos años?
- Preguntas que evalúen los conocimientos de los encuestados de forma abrupta y lleven al encuestado a sentirse incómodo. Ej. ¿Sabe cuáles son las funciones del Jefe del Estado? ¿De cuántos ministros se compone el gobierno actual?
- Preguntas demasiado generales. Ej. ¿Le gusta el presidente?
- Preguntas redactadas de forma negativa. Ej. ¿No ha visitado a su médico en el último mes?
- Preguntas que afecten la sensibilidad de los entrevistados. Ej. ¿Es usted racista?
- Preguntas que contengan algunas de las opciones que están en las respuestas. Ej. ¿Qué medio de transporte como bus o taxi usa para ir a su trabajo?
- Palabras abstractas o polisémicas que pueden ser interpretadas de diversa manera. Ej. ¿Cree que el país se ha desarrollado en el último año?
- Palabras y frases ambiguas. Ej. ¿Cuánto ha progresado en matemáticas?
- Juicios implícitos Ej. Como usted ya sabe...
- Juicios de valor Ej. Aunque la religión se ha convertido en un asunto controvertido en nuestro país, ¿cada cuánto asiste a servicios religiosos?

Hay varios **tipos de preguntas** de cuestionario

Figura 4. Tipos de preguntas de cuestionario



Fuente: elaboración propia

Puede encontrarse más detalle sobre el tipo de preguntas de un cuestionario en e-encuesta.com (2015).

Las **preguntas de tipo Likert** son aquellas donde se pregunta en forma de escala por el grado de conformidad del encuestado con una proposición. Este tipo de preguntas deben su nombre al psicólogo Rensis Likert que en 1932 quería medir las actitudes evitando las preguntas dicotómicas de Sí o No. Por ejemplo:

¿En qué grado está de acuerdo o en desacuerdo con la política de extradición de narcotraficantes a Estados Unidos?

- 1- Totalmente de acuerdo
- 2- De acuerdo
- 3- Ni de acuerdo ni en desacuerdo
- 4- En desacuerdo
- 5- Totalmente en desacuerdo
- 0- No sabe / No contesta

Las categorías de repuesta pueden comprender de 1 a 5, de 1 a 10, o de 1 a 4 según decida el investigador. Si se usa una escala con un total de categorías que sea número par, como de 1 a 4, se elimina la categoría intermedia (ni de acuerdo ni en desacuerdo) y obliga al encuestado a tomar una decisión a favor o en contra de algo. Sea como fuere, cuando se quiera preguntar por cuestiones de opinión, conformidad o agrado se recomienda el uso de preguntas Likert.

### 3. Redacción de las categorías de respuesta

Las categorías u opciones de respuesta acompañan a cada pregunta cerrada y deben cumplir dos características. Las categorías de respuesta deben ser exhaustivas y mutuamente excluyentes.

- Exhaustivas: se debe ofrecer todo el abanico de respuestas posibles para que el encuestado se sitúe en alguna de ellas. Las categorías de repuesta de la siguiente pregunta no cumplen el criterio de exhaustividad.

*"Independientemente del lugar donde accede a Internet, ¿para qué lo usa Ud.?"*

1. *Para usar el correo electrónico*
2. *Para buscar información*
3. *Para entretenerse*
4. *Para trabajar*
5. *Para hacer trámites*
0. *NS/NC*

En esta pregunta faltan ítems o categorías de respuesta, como “para hacer compras” o “para estudiar”. Además, como son múltiples y muy diversos los usos de Internet, es conveniente colocar la categoría “Otros”. Al colocar la categoría “Otros” en una pregunta donde las opciones de respuesta son casi infinitas, cumplimos el criterio de exhaustividad. La forma correcta de redactar las categorías de respuestas sería:

*Independientemente del lugar donde accede a Internet, ¿para qué lo usa Ud.?"*

1. *Para usar el correo electrónico*
2. *Para buscar información*
3. *Para entretenerse*

4. *Para trabajar*
5. *Para estudiar*
6. *Para hacer trámites*
7. *Para hacer compras*
8. *Otros*
0. *NS/NC*

- Mutuamente excluyentes: las categorías de respuesta no se pueden solapar ya que el encuestado se debe situar en una, y solo en una de las opciones de respuesta. Las categorías de respuesta de la siguiente pregunta no cumplen el criterio de ser mutuamente excluyentes.

*¿Con qué frecuencia habla usted de política con sus amigos?*

1. *Muy frecuentemente*
2. *Frecuentemente*
3. *Bastante*
4. *Casi nunca*
5. *Nunca*
6. *No hablo de política*
7. *NS/NC*

Las categorías “2. Frecuentemente” y “3. Bastante” son muy similares y se solapan, por lo que es conveniente descartar una de ellas. Además, si usted no habla nunca de política con sus amigos podría situarse en las categorías “5. Nunca” y “6. No hablo de política”. Por tanto, también conviene eliminar una de ellas. La pregunta y sus categorías están correctamente redactadas de la siguiente manera:

*¿Con qué frecuencia habla usted de política con sus amigos?*

1. *Muy frecuentemente*
2. *Frecuentemente*
3. *Casi nunca*
4. *Nunca*
5. *NS/NC*

#### **4. Secciones de un cuestionario**

Se recomienda seguir la siguiente organización para presentar el cuestionario y ordenar las preguntas por bloques temáticos:

**Presentación** – todo cuestionario debe contener un mensaje de bienvenida donde se presente la encuesta, la finalidad u objetivo del estudio, el uso posterior de los datos y ofrecer un email o teléfono de contacto por si el encuestado quiere contactar al responsable del estudio. Además, si no se pregunta por el nombre del encuestado, lo cual es recomendable, debe explicitarse en este mensaje que se garantiza el anonimato del encuestado. Esta presentación debe ser muy corta y motivar al encuestado a que conteste las preguntas.

**Preguntas iniciales** – aquellas importantes para asegurarnos que el encuestado las responde en el momento que tiene mayor atención, pero que no sean invasivas de la intimidad del encuestado.

**Preguntas críticas** – aquellas que sean más problemáticas o invasivas.

Preguntas sociodemográficas – son aquellas preguntas sobre el perfil social y demográfico del encuestado: género, año de nacimiento,<sup>2</sup> lugar de residencia, nivel educativo, nivel de ingresos, profesión. El investigador decide cuáles pueden ser relevantes para su estudio, aunque se recomienda incluirlas todas ya que son explicativas de las respuestas u opiniones anteriormente recogidas.

Se recomienda que el cuestionario no sea excesivamente largo para evitar faltas de atención del encuestado o abandono prematuro, sobre todo si el cuestionario se distribuye online.

## **5. Codificación**

Toda pregunta y categoría de respuesta debe ir asociada a un código, se recomienda que se usen letras numeradas para las preguntas y números para las categorías de respuesta ya que de esta forma se guardan en los softwares de análisis de datos.

## **6. Validación de la encuesta**

Toda encuesta nueva que hayamos diseñada debe ser validada. El proceso de validación de una encuesta consta de dos fases: externa e interna.

La validación externa es la revisión por parte de expertos en el tema de la encuesta y en metodología de investigación para comprobar si las preguntas están midiendo los indicadores, si están correctamente formuladas y si el cuestionario presenta algún error o posibles mejoras.

La validación interna es una prueba de la encuesta con un pequeño grupo de personas. Se le aplica el cuestionario a una parte pequeña de la muestra para saber si han entendido todas las preguntas, si están representados en las categorías de respuesta, si la referencia de tiempo es clara, y si detectan algún error o posible mejora.

La población de estudio y la muestra a la que aplicaremos la encuesta es un aspecto esencial para la redacción de las preguntas. Qué procedimiento y tamaño de muestreo se ha de usar lo veremos en el siguiente capítulo.

## **Bibliografía sugerida**

Cea d'Ancona, M.A. (1999) Metodología Cuantitativa: Estrategias y Técnicas de Investigación Social. Capítulo 7. Madrid: Síntesis

[http://www.trabajosocial.unlp.edu.ar/uploads/docs/metodologia\\_cuantitativa\\_estrategias\\_y\\_tecnicas\\_de\\_investigacion\\_social\\_cea\\_d\\_ancona.pdf](http://www.trabajosocial.unlp.edu.ar/uploads/docs/metodologia_cuantitativa_estrategias_y_tecnicas_de_investigacion_social_cea_d_ancona.pdf)

Díaz de Rada, V. (2001) *Diseño y elaboración de cuestionarios para la investigación comercial*. Madrid: ESIC Editorial

---

<sup>2</sup> Para preguntar por la edad del encuestado es mejor preguntar por el año de nacimiento antes que interrogar directamente por cuántos años tiene la persona, esto aumenta la tasa de respuestas ya que a muchas personas no les gusta confesar su edad y menos a un desconocido.

e-encuesta.com (2015) Tipos de pregunta para una encuesta. *e-encuesta.com*.  
Disponible en: <http://www.e-encuesta.com/blog/2015/tipos-de-pregunta-en-la-encuesta/>  
(último acceso 17 de abril de 2017).

### **Ejercicio 1**

A partir de los indicadores identificados en la fase anterior, construya preguntas de cuestionario para recoger información sobre cada uno de ellos. Tenga en cuenta todos los aspectos señalados en esta lección antes de redactar sus preguntas de cuestionario.

### **Ejercicio 2**

Revise la encuesta Latinobarómetro y seleccione aquellas preguntas que le servirían en su cuestionario. Por ejemplo, las preguntas sociodemográficas. Adáptelas si cree conveniente

### **Ejercicio 3**

A partir de las preguntas redactas en el ejercicio 1 y las adaptadas o copiadas en el ejercicio 2, construya el cuestionario en Google Drive, en la opción Formularios. En Youtube hay diversos tutoriales sobre cómo construir un cuestionario en Google Drive. Una vez terminado, distribúyalo entre al menos 5 de sus contactos para realizar una prueba piloto de validación interna, y posteriormente entre 2 profesores para realizar la validación externa.



## **7. Cómo seleccionar y calcular muestras de una población**

### **Qué es una muestra y qué es una población o universo**

El universo o población de estudio es el conjunto de actores (personas, organizaciones, países...) que comparten alguna/s característica/s y que son aquellas unidades de análisis con los que se responderá la pregunta de investigación. La muestra es un número reducido de actores tomados de esa población. Estudiamos muestras debido a la imposibilidad de estudiar toda la población o universo.

El objetivo principal del muestreo es la posibilidad de generalizar, es decir, extraer conclusiones generales basadas solo en el estudio de unos cuantos casos. Cómo se obtiene una muestra representativa, cómo se seleccionan los casos de la muestra y qué tamaño debe tener una muestra es importante ya que malos muestreos llevan a malas conclusiones, y por tanto a malas investigaciones.

Hay tres asuntos que debemos preguntarnos a la hora de extraer una muestra: 1) representatividad de la muestra, 2) tipos de muestreo, 3) tamaño de la muestra

#### **1. Representatividad de la muestra**

##### **¿Queremos una muestra representativa?**

Si queremos generalizar, es decir, señalar que lo estudiado con una muestra se puede atribuir a toda la población, necesitamos muestras representativas. Tenemos una muestra representativa cuando las principales características de la población están presentes en proporción en la muestra. Es el investigador, en función de su pregunta de investigación e hipótesis, el que debe decidir qué variables son tomadas en cuenta para realizar el muestreo.

La proporción puede ser extraída equitativamente del número de categorías de la variable, o en proporción equivalente a como las diferentes categorías están en la población. Por ejemplo, si consideramos que es importante en nuestra investigación la variable género y sus dos categorías son hombre y mujer, la muestra deberá contener tanto hombres como mujeres. Puede ser 50% hombres y 50% mujeres (criterio equitativo), o bien la proporción equivalente que haya en la población, si hay 800 hombres y 200 mujeres en el universo o población, la muestra deberá contener 80% de hombres y 20% de mujeres (criterio equivalente). La decisión final de si serán 50% hombres y 50% mujeres, o 80% hombres y 20% mujeres, la toma el investigador basándose en los objetivos de la investigación y las hipótesis a comprobar.

#### **2. Tipos de muestras**

##### **¿Qué tipo de procedimiento seguiremos para seleccionar los actores de una muestra?**

Hay dos grandes estrategias: muestreo no-aleatorio y muestreo aleatorio. También se suelen denominar muestreo no-probabilístico y probabilístico.

a. Muestreo no-aleatorio: el investigador deliberadamente elige los casos y no el azar. Hay varios subtipos:

- i. Por conveniencia o accidental: el investigador elige aquellos actores que estén a su alrededor y disponibles. Por ejemplo: encuestar en una calle a los transeúntes.
- ii. Por cuotas: el investigador elige los actores teniendo en cuenta que haya actores de las diversas categorías de una variable. Por ejemplo: encuestar a un grupo de hombres y a un grupo de mujeres proporcionalmente, es decir, con cuotas de ambas categorías de la variable género.
- iii. Intencional o de juicio: el investigador selecciona los casos siguiendo su criterio o el de expertos. Por ejemplo, encuestar a personajes claves siguiendo el juicio de un experto en el tema.
- iv. Bola de nieve: un actor entrevistado pone en contacto al investigador con otros posibles actores a entrevistar. Útil para investigar poblaciones clandestinas o de difícil acceso por el investigador.
- v. Voluntariado: son los sujetos o actores de estudio los que acuden para ser investigados ante una convocatoria o llamado del investigador. Usado en investigaciones experimentales donde se presentan candidatos voluntariamente para la investigación.

b. Muestreo aleatorio: el azar selecciona los actores. Todos los actores de la población deben tener las mismas posibilidades de ser elegidos. Hay varios subtipos:

- i. Aleatorio simple: se realiza un “sorteo” entre todos los actores. Se necesita el listado de toda la población. Se enumeran todos los actores y se elige al azar el número de actores según el tamaño de la muestra.
- ii. Aleatorio sistemático: teniendo el listado enumerado de la población se van dando “saltos” en la lista siguiendo un intervalo fijo para seleccionar a los actores que compondrán la muestra. Pasos: a) se dispone del listado de la población de p.ej. 10.000 actores; b) se divide el tamaño de la población por el tamaño de la muestra y así obtenemos el intervalo, p.ej.  $10.000/200 = 50$ , es decir, daremos saltos de 50 en 50 en el listado de la población; c) escogemos un actor de la lista al azar y a partir de él contamos el número del intervalo (50) para escoger al siguiente, y a partir de ese, contando 50 más obtenemos el siguiente, así hasta alcanzar el tamaño de la muestra. Útil en encuestas telefónicas usando el directorio o guía telefónica.
- iii. Aleatorio estratificado: se necesita también el listado de la población y se realiza un “sorteo” de los actores, pero teniendo en cuenta recoger actores de categorías de una variable relevante en la investigación. Por ejemplo: seleccionar aleatoriamente hombres y mujeres en proporción equivalente a la población.
- iv. Aleatorio por conglomerados: en este caso se realiza un sorteo de los lugares, organizaciones o conglomerados a los que pertenecen los actores y posteriormente se escogen los actores que pertenecen a estas organizaciones. Por ejemplo: para entrevistar a los estudiantes de un país debido a que no existe un registro centralizado de todos los estudiantes del país, se seleccionan al azar universidades y se entrevistan a los estudiantes de las universidades elegidas. Si se sortean posteriormente subdivisiones de estos conglomerados u organizaciones se trata de un muestreo por conglomerados polietápico (varias etapas). Siguiendo con el ejemplo anterior, como el número de estudiantes de las universidades es elevado,

en una segunda etapa, se sortean facultades de las universidades seleccionadas. De las facultades elegidas se seleccionan al azar aulas en esta tercera etapa y ahí se entrevistan a los estudiantes.

Lo más habitual es que se combinen los diferentes tipos de muestreo. El más usado en investigaciones sociológicas a nivel macro suele ser el conglomerado polietápico estratificado, aunque todos son igual de válidos y legítimos. Es importante justificar por qué se elige un procedimiento u otro. Se recomienda revisar investigaciones empíricas para ver cómo justifican el uso de una estrategia de muestreo.

### **3. Tamaño de la muestra**

#### **¿Qué tamaño debe tener la muestra?**

Si queremos que la muestra sea representativa seguiremos un proceso de cálculo del tamaño teniendo en cuenta 4 factores:

- a. **Tamaño de la población:** cuanto más grande es una población, mayor será el tamaño de la muestra. Aunque llega un momento que no por aumentar más el tamaño de la muestra más representativo serán los resultados.
- b. **Heterogeneidad:** hace referencia al nivel de dispersión de la población en alguna variable relevante en la investigación. Cuanto más heterogénea sea una población, mayor será el tamaño de la muestra. Por ejemplo, si sabemos que en una población a encuestar el 70% de las personas tienen un perfil homogéneo, el porcentaje de heterogeneidad será 70% o 0.70. Si no hay ninguna variable clave en nuestra investigación, la heterogeneidad será 50% o 0.5 que representa la mayor heterogeneidad. Suele ser común dejar el nivel de 50% ya que es una medida estándar.
- c. **Error muestral o margen de error:** hace referencia a cuánto varían los resultados de la muestra respecto del universo o población. Equivale a un rango de valores en el que se encuentra el resultado de la población. Por ejemplo, si la media de veces que va al cine al mes una muestra es de 2,5 y la media de veces que va al cine al mes la población es 1, el error muestral es la diferencia entre 2,5 y 1. El error muestral lo escoge el investigador antes de realizar el estudio o investigación y se suele elegir de forma estándar. En investigaciones sociológicas se usa comúnmente un error muestral del 3% o 5%. Cuanto menor sea el error muestral, más grande será el tamaño de la muestra ya que para ser más precisos (menos error) deberemos estudiar más actores de la población (más tamaño de la muestra).
- d. **Nivel de confianza:** es la probabilidad que el resultado obtenido se encuentre dentro del intervalo de confianza. Se suelen utilizar niveles de confianza del 95% en investigaciones sociológicas. Es un poco complejo de entender pero con un ejemplo se clarifica. Estudiamos una muestra usando un nivel de confianza del 95% y obtenemos que la media de consumo de cigarrillos por día es 5,5 y la desviación estándar es 1,5. El intervalo de confianza del 95% se expresa como la media  $\pm 1,96$  desviaciones estándar, es decir, el intervalo irá entre aproximadamente 8,5 ( $5,5 + 1,96$ ) y 2,5 ( $5,5 - 1,96$ ). Es decir, hay un 95% de posibilidades que la media de cigarrillos consumidos al día en toda la población se encuentre entre 8,5 y 2,5. Si usamos niveles de confianza mayores para ser más precisos como el 99%, el tamaño de la muestra será mayor.

La fórmula para calcular el tamaño de la muestra es:

$$n = \frac{N * Z^2 * p(1 - p)}{(N - 1) * e^2 + Z^2 * p(1 - p)}$$

n: tamaño de la muestra  
N: tamaño de la población  
Z: nivel de confianza  
e: error muestral  
p: heterogeneidad

Si el tamaño de la población es muy grande, más de 100.000 personas, la fórmula puede simplificarse así:

$$n = \frac{Z^2 * p(1 - p)}{e^2}$$

Hoy en día existen calculadoras online para saber el tamaño de la muestra conveniente teniendo en cuenta estos factores. Es fácil usarlos y permite jugar con diferentes niveles para saber el tamaño de la muestra más adecuado. Incluso permite calcular el error muestral si ya tenemos elegido o definido el tamaño de la muestra.

<https://www.netquest.com/es/panel/calculadora-muestras/calculadoras-estadisticas>

### **Ejercicios**

A partir de su pregunta de investigación y población de estudio con la que desea responderla, responda a las siguientes preguntas:

- ¿Optaría por analizar una muestra representativa o no-representativa? ¿Qué ventajas tendría en un caso y en otro?
- ¿Qué procedimiento de muestreo cree que es el más adecuado? ¿Por qué? Tenga en cuenta el posible acceso a la población de estudio, los recursos disponibles, el conocimiento completo de la población y otros factores para justificar su respuesta
- Calcule cuál sería el tamaño de muestra a partir de las condiciones de su estudio.

## 8. Cómo procesar y guardar los datos recolectados

Procesar los datos significa ordenar, evaluar y guardar el material recolectado en matrices de datos para posteriormente analizarlos. Es decir, las respuestas de una persona se sistematizan en una base de datos, se ordenan y guardan para que puedan ser analizadas. A partir del procesamiento y análisis de los datos se usan los siguientes términos:

- **Caso:** son los actores o unidades de análisis del estudio. Pueden ser personas, cosas, organizaciones, países. Ejemplo de caso: las personas que han respondido un cuestionario. Otro ejemplo de caso: los países que han tenido guerra civil.
- **Variable:** es una propiedad o característica de los actores que adopta diversos valores. Son los indicadores que medimos de los actores, las preguntas que se les hace a los actores. Ejemplos de variables: género, nivel educativo, estatura, peso, beneficios económicos al final de año, extensión geográfica, índice GINI, número de protestas.
- **Categorías de respuesta o valores:** son las opciones de respuesta o valores de una variable. Ejemplo: masculino, femenino (para la variable género); sin estudios, estudios primarios, secundarios y universitarios (para la variable nivel educativo); 180 cm, 181 cm (para la variable estatura); 83 kg, (para la variable peso); 1.000 euros (para la variable beneficios económicos al final del año); 600 km<sup>2</sup> (para la variable extensión geográfica); 0,23 y 0,53 (para la variable índice GINI); 0, 1, 2, 3... (para la variable número de protestas).

Los datos recolectados ya sean a través de fuentes primarias (p.ej. encuesta hecha por el propio investigador) o de fuentes secundarias (datos producidos por otro grupo de investigadores y recolectados de internet) deben ser guardados en matrices de datos. Estas matrices de datos incluyen dos hojas. La hoja de variables y la hoja de los casos.

### La hoja de las variables

Se incluye información sobre las variables y las categorías de respuesta. El formato depende del programa informático, pero la información que contiene es casi siempre: código de la variable, nombre de la variable, tipo de datos, valores o categorías de respuesta y sus códigos, tipo de variable.

### La hoja de los casos

En la hoja de casos se incluyen los valores (o respuestas) de las variables para cada caso: en las filas se coloca la lista de casos que son incluidos en la muestra, en las columnas se colocan las variables (o preguntas del cuestionario), en las celdas irán los valores o respuestas. En casi todos los programas informáticos el formato es similar.

Tabla 3. Ejemplo de hoja de variables

Código de la variable	Nombre de la variable	Tipo de datos	Valores o categorías de respuesta y sus códigos	Tipo de variable	...
A1	Género	Numérico	1- Masculino 2- Femenino	Nominal	
A2	Nivel Educativo	Numérico	1- Sin estudios 2- Primarios 3- Secundarios 4- Universitarios 0- No sabe / No contesta	Ordinal	
A3	Año de nacimiento	Numérico		Escalar	
A4	Votar en las últimas elecciones	Numérico	1- Sí 2- No 0- No sabe / No contesta	Nominal	

Tabla 4. Ejemplo de hoja de casos

Número de caso	Variable A1. Género	Variable A2. Nivel educativo	Variable A3. Año de nacimiento	Variable A4. Votar en las últimas elecciones	...
Caso 1	1. Masculino	Universitarios	1981	Sí	
Caso 2	Femenino	Universitarios	1976	Sí	
Caso 3	Femenino	Primarios	1956	No	
Caso 4	Masculino	Secundarios		Sí	
Caso 5					
Caso 6					

Es recomendable que los valores o respuestas sean codificados en números, por tanto, esta hoja de casos sería:

Número de caso	Variable A1. Género	Variable A2. Nivel educativo	Variable A3. Año de nacimiento	Variable A4. Votar en las últimas elecciones	...
Caso 1	1	4	1981	1	
Caso 2	2	4	1976	1	
Caso 3	2	2	1956	2	
Caso 4	1	3		1	
Caso 5					
Caso 6					

### Casos perdidos

Cuando un caso (p.ej. una persona) no responde una pregunta de cuestionario (variable) se deja la casilla en blanco. A este caso se le denomina caso perdido (en inglés *missing case*). En general, si nuestra muestra es grande, estos casos perdidos son extraídos del análisis. En muestras pequeñas el valor del caso perdido puede ser sustituido por el valor promedio de la variable. Hay diversas estrategias para llevar a cabo investigaciones con alto número de casos perdidos, véase Mallou et al. (1998)

### Bibliografía sugerida

Mallou, J. V., Tobío, T. B., Carreira, A. G., Boubeta, A. R., & Fernández, X. G. V. (1998). Estimación de la respuesta de los «no sabe/no contesta» en los estudios de intención de voto. *REIS*, 83, 269-287. <https://doi.org/10.2307/40184128>

### Ejercicio

Cree en Excel una hoja de variables y una hoja de casos con las preguntas de cuestionario que redactó en ejercicios anteriores, y con las respuestas obtenidas en la prueba piloto de validación. Complete la hoja de casos con las respuestas que obtenga de sus compañeros o de una muestra de su población.

## 9. Programas informáticos para análisis estadístico de datos cuantitativos

El análisis estadístico y la investigación cuantitativa no serían tan populares sin los programas informáticos. La función de los programas informáticos en la investigación es ejecutar análisis con cantidades de datos considerables, por lo que son imprescindibles en la investigación de datos cuantitativos.

Los principales programas informáticos para el análisis estadístico de datos cuantitativos y usados en ciencias sociales son: Excel, SPSS, PSPP, Stata y R, entre muchos otros. Cada uno tiene sus ventajas y desventajas, y la decisión de usar uno u otro depende de los recursos disponibles y la familiaridad del investigador con programación.

En este capítulo presentamos el software PSPP que fue creado a imagen y semejanza de SPSS pero con código abierto. SPSS es el software más popular en ciencias sociales debido a su alta usabilidad y enorme capacidad de análisis, pero con un coste de licencia elevado, solo al acceso de universidades o empresas. Por ello, PSPP fue creado mediante software libre que permite la distribución y acceso libre. Véase que PSPP son las siglas invertidas de SPSS.

Descarga del software PSPP: <https://www.gnu.org/software/pspp/>

Manual de instalación de PSPP en español:

[http://www2.uned.es/socioestadistica/Practicas\\_%20ejercicios\\_guia/Practica2.pdf](http://www2.uned.es/socioestadistica/Practicas_%20ejercicios_guia/Practica2.pdf)

Tutorial de procesamiento de datos en PSPP:

[http://www2.uned.es/psiped-psicologia-social/ecodilema/acceso/orientaciones\\_analisis.pdf](http://www2.uned.es/psiped-psicologia-social/ecodilema/acceso/orientaciones_analisis.pdf)

En Youtube existen decenas de tutoriales de cada una de las funciones y análisis estadísticos de PSPP. Puede consultarlos si desea conocer más sobre este software libre.

El programa informático ejecuta las órdenes del investigador, por lo que se debe tener claro qué análisis corresponde en cada momento según el tipo y número de variables, como veremos en el próximo capítulo.

### Bibliografía sugerida

Salmerón Gómez, Romelio (2015) (Mini)Manual de PSPP, alternativa libre a SPSS. Disponible en: <https://softtcm.files.wordpress.com/2014/04/pspp.pdf>

### Ejercicio

Descargue las bases de datos de la Encuesta Mundial de Valores y Latinobarómetro en formato ASCII, SPSS, Stata o R. Revise cómo está guardada la información de los casos y las variables.



## 10. Análisis de datos: tipos de análisis estadísticos

La estadística es una ciencia de las matemáticas que se usa en la investigación para analizar datos cuantitativos.

El análisis de los datos depende de dos grandes factores:

- A. el tipo de variables que analizamos
- B. el número de variables que analizamos simultáneamente

Las variables equivalen a las preguntas del cuestionario o los indicadores identificados

### A. Tipos de variables

Hay 3 grandes tipos de variables según las categorías de respuesta:

- **Nominales:** son aquellas variables cuyas categorías de respuesta no tienen un orden preestablecido ni jerarquía interna. Es decir, las categorías de respuesta pueden variar de orden y mantienen su sentido. Ejemplos:
  - Variable Género: 1- Masculino, 2-Femenino
  - Variable Estado civil: 1-Soltero, 2-Casado, 3-Vive en pareja, 4-Divorciado, 5-Viudo, 6-Otros
- **Ordinales:** son aquellas variables cuyas categorías de respuesta sí tienen un orden preestablecido o jerarquía interna. Las categorías de respuesta deben mantener un orden. Ejemplos:
  - Variable Nivel educativo: 1-Sin estudios, 2-Estudios primarios, 3-Estudios secundarios, 4-Estudios universitarios
  - Variable Grado de conformidad con el presidente: 1-Totalmente de acuerdo, 2-De acuerdo, 3-Ni de acuerdo ni en desacuerdo, 4-En desacuerdo, 5-Totalmente en desacuerdo
  - Variable grupo de edad: 1-Menos de 18 años, 2-Entre 18 y 35 años, 3-Entre 36 y 50 años, 4-Entre 51 y 65 años, 5-Más de 65 años
- **Escalar:** son aquellas variables cuyas categorías de respuestas sí tienen un orden preestablecido o jerarquía interna, y el espacio entre una categoría y otra es el mismo.
  - Variable Año de Nacimiento (pregunta abierta):
  - Variable Número de veces que ha asistido al cine el último mes (pregunta abierta):

Cuando la edad se pregunta de forma abierta es una variable escalar, cuando se pregunta por intervalos es una variable ordinal, ya que, al preguntarlo de forma abierta, el espacio que hay entre 18 y 19 años (365 días) es el mismo que hay entre 35 y 36 años (también 365 días), y entre 65 y 66 años (también 365 días). Todas las variables o preguntas que hacen referencia a cantidades y no están preguntadas en forma de intervalos son variables escalares.

### B. Número de variables a analizar

Según el número de variables a analizar simultáneamente, hay 3 grandes tipos de análisis:

- **Análisis descriptivo univariado:** se analiza una sola variable, y este tipo de análisis tiene una finalidad netamente descriptiva, es decir, exponer los resultados de una variable.

- **Análisis bivariado:** se analizan dos variables, y este tipo de análisis tiene una finalidad de comprobar hipótesis relacionales (causales o asociativas), es decir, se comprueba si dos variables están relacionadas o no.
- **Análisis multivariado (o multivariante):** se analizan más de dos variables, y este tipo de análisis tiene una finalidad de comprobar hipótesis relacionales (causales o asociativas), es decir, se comprueba si varias variables están relacionadas entre sí.

### Técnicas estadísticas

Las principales técnicas de análisis estadístico en cada uno de los tipos de análisis son:

Tabla 5. Técnicas estadísticas principales

Tipo de análisis	Técnicas estadísticas
Análisis univariado	<ul style="list-style-type: none"> <li>▪ Tablas de frecuencias</li> <li>▪ Media, desviación típica</li> </ul>
Análisis bivariado	<ul style="list-style-type: none"> <li>▪ Tablas de contingencia</li> <li>▪ Correlaciones bivariadas</li> <li>▪ ANOVA de un factor</li> </ul>
Análisis multivariado	<ul style="list-style-type: none"> <li>▪ Regresión lineal múltiple</li> <li>▪ Regresión logística (binaria y múltiple)</li> <li>▪ Modelos logit y probit</li> <li>▪ Análisis discriminante</li> <li>▪ Análisis factorial</li> <li>▪ Análisis de clústers</li> <li>▪ Escalado multidimensional</li> <li>▪ ANOVA de dos factores</li> <li>▪ MANOVA</li> <li>▪ Ecuaciones estructurales</li> </ul>

Fuente: elaboración propia

En los siguientes capítulos se explicará cómo usar las técnicas estadísticas para llevar a cabo estos tipos de análisis. La estadística permite describir y hacer inferencias. La estadística descriptiva busca sintetizar y visualizar los resultados de estudio. La estadística inferencial buscar sacar conclusiones o predicciones a partir del análisis de los datos obtenidos en un estudio.

### Bibliografía sugerida

Estadística para torpes. Disponible en:

<https://unidadinvestigacionhvn.files.wordpress.com/2010/11/estadistica-para-torpes.pdf>

### **Ejercicio**

Seleccione diversos artículos de investigación relacionados con su pregunta de investigación o su tema de interés. Compruebe qué técnicas estadísticas emplean para comprobar las hipótesis o responder a la pregunta de investigación.

## 11. Análisis univariado descriptivo

Los estadísticos usados para analizar una sola variable con la finalidad de describir los datos recolectados dependen del tipo variable: nominal, ordinal o escalar.

Si la variable que deseamos analizar es *nominal*, se usará una tabla de frecuencias y un gráfico circular o de barras.

Si la variable que deseamos analizar es *ordinal*, se usará una tabla de frecuencias y un gráfico de barras

Si variable que deseamos analizar es *escalar u ordinal de más de cinco categorías*, se usarán los estadísticos de media (promedio) y desviación estándar (típica), y un histograma como gráfico. Ojo, no se recomienda analizar los datos mediante una tabla de frecuencias cuando la variable es de tipo escalar.

### Tabla de frecuencia

La tabla de frecuencias es un análisis donde se expresan las respuestas de los datos recolectados en la cantidad de veces que ocurren (frecuencia absoluta) y el porcentaje que representan estas respuestas (frecuencia relativa). La frecuencia relativa (o porcentaje de respuestas) se calcula dividiendo el número de respuestas de una categoría entre el total de casos.

Ejemplo de análisis descriptivo univariado mediante tabla de frecuencias y gráfico de barras

Variable: "*¿Alguna vez en su vida ha usado Ud. correo electrónico o se ha conectado a Internet?*"

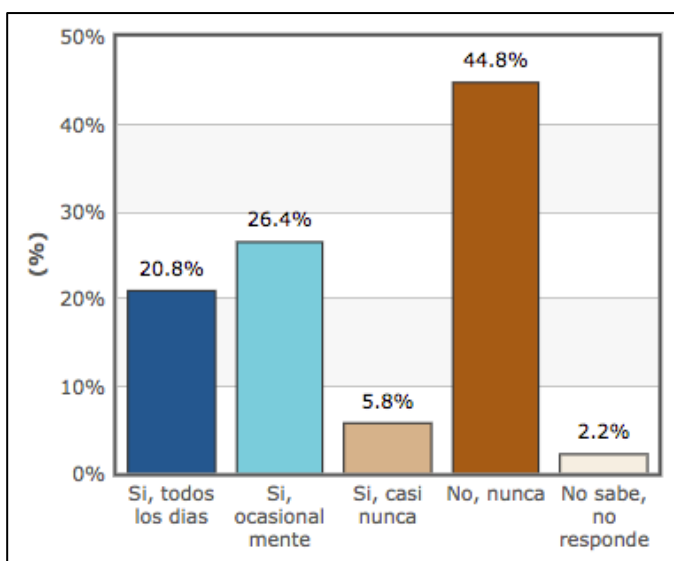
Casos: personas habitantes en Perú

Tabla 6. Ejemplo de tabla de frecuencias

	Frecuencia absoluta (número de casos)	Frecuencia relativa (porcentaje)
Sí, todos los días	250	20,8%
Sí, ocasionalmente	317	26,4%
Sí, casi nunca	70	5,8%
No, nunca	537	44,8%
No sabe, no contesta	26	2,2%
N (total de casos)	1200	100%

Fuente: elaboración propia a partir de Latinobarómetro (2015)

Figura 5. Ejemplo de gráfico de barras



Fuente: elaboración propia a partir de Latinobarómetro (2015)

Todo análisis debe ir acompañado de un texto donde se interpreten los resultados, es decir, se destaquen los principales valores o categorías de respuesta y se haga una lectura sobre ello. Ejemplo:

La mayoría de encuestados en Perú (44,8%) nunca ha usado correo electrónico o se ha conectado a Internet en 2015. Sorprende este resultado en la era digital, lo que manifiesta la baja extensión de las tecnologías de la información en Perú.

### **Media (promedio)**

La media (también denominada popularmente promedio o en lenguaje estadístico, media aritmética) es el valor que representa y sintetiza un conjunto de datos. Es una medida de la tendencia central de una variable. Se calcula a partir de la suma de los valores de las respuestas dividido por el número total de casos.

### **Desviación estándar**

La desviación estándar (también denominada desviación típica) es el valor que mide la dispersión de valores o respuestas de una variable. Es el promedio de las distancias de los valores de cada caso respecto a la media. Es decir, se calcula una media de cuánto se alejan las respuestas de cada actor respecto a la media. Cuanto más alta sea la desviación estándar más dispersas son las respuestas obtenidas, es decir, más heterogénea es la opinión o comportamiento de los actores analizados.

La desviación estándar al cuadrado es la varianza. Tanto la desviación estándar como la varianza son usados ampliamente en los análisis multivaridos.

El histograma es el gráfico usado para visualizar los resultados de una variable escalar, y se caracteriza por incluir la curva de normalidad. La curva normal, también conocida como campana de Gauss, es una representación de la distribución de los datos siguiendo

Ejemplo de análisis descriptivo univariado mediante media y desviación estándar.

Variable: *Con una escala de 1 a 10, le pedimos evaluar cuán democrático su país. El “1” quiere decir que “su país no es democrático” y el “10” quiere decir que “su país es totalmente democrático” ¿Dónde pondría Ud. a su país?*

Se realizó esta pregunta en tres países diferentes. Debido a que es una variable ordinal de más de 5 categorías se asemeja a una variable escalar y se calcula la media y desviación estándar.

Resultados:

Tabla 7. Ejemplo de tabla comparación de medias y desviación estándar

País	Media	Desviación estándar
Perú	5,26	2,05
Chile	5,78	1,98
México	5,00	2,46

Fuente: elaboración propia a partir de Latinobarómetro (2015)

La interpretación de estos resultados sería:

Según las opiniones de los encuestados en sus respectivos países, el país más democrático es Chile (media = 5,78), seguido de Perú (media = 5,26), y el menos democrático de los tres analizados es México (media = 5,00). Aunque en México es donde más divergen las opiniones (desviación estándar = 2,46).

Para decir que una media y una desviación estándar son altas, media o bajas debemos comparar los resultados entre los grupos de otra variable. En el ejemplo anterior, hemos comparados los resultados de la pregunta sobre “qué tan democrático es su país” en función del país. Para saber si estas diferencias son significativas llevaremos a cabo técnicas más complejas como ANOVA de un factor, que veremos en capítulos siguientes.

En conclusión, el tipo de análisis para describir una variable depende del tipo de variable

Tabla 8. Tipo de variables y análisis descriptivo univariado

Variabes	Tabla de frecuencias	Medida de tendencia central	Medidas de dispersión	Gráficos
Nominal	Sí	<i>Moda*</i>	<i>Ratio de variación*</i>	Sectores o barras
Ordinal	Sí	<i>Mediana*</i>	<i>Rango*</i>	Barras
Escalar **	No	Media	Desviación típica	Histograma

\* Tienen un uso limitado, por lo que no los presentamos ya que no aportan mucho a la descripción de la variable. \*\* Las ordinales de más de 5 categorías pueden ser tratadas como variables escalares.

Fuente: elaboración propia a partir de Latinobarómetro (2015)

### Ejercicio

Revise la Encuesta Mundial de Valores en su opción de “Online Analysis”. Escoja una sección o tema y analice todas las variables de esa sección. Presente los resultados del análisis descriptivo univariado para cada variable. Tenga en cuenta el tipo de variable para seleccionar el tipo de análisis (tabla de frecuencias, media y desviación estándar) y el tipo de gráficos (sectores, barras o histograma).

## 12. Cómo hacer y analizar tablas de contingencia

Las tablas de contingencia (también llamadas a veces tablas dinámicas, tablas cruzadas, tablas de control o *crosstabs* en inglés) son posiblemente la técnica estadística más utilizada en análisis de datos. Las tablas de contingencia son una forma de presentar los datos de dos variables, y pueden ser usadas para analizar la asociación entre dos variables mediante las frecuencias relativas (porcentajes).

### Condiciones para usar tablas de contingencia:

- Solo se pueden relacionar dos variables, por ello son una técnica de análisis bivariado.<sup>3</sup>
- Esta técnica de análisis se usa especialmente con variables nominales y ordinales. Las variables nominales son las que no tienen orden interno establecido (p.ej. género o estado civil), y las variables ordinales son aquellas que sí tienen un orden interno establecido y el paso de una categoría a otra no es igual (p.ej. nivel educativo (sin estudios-primarios-secundarios-universitarios) o interés en la política (mucho interés-algo de interés-no muy interesado-nada interesado)). Las tablas de contingencia no se usan para analizar relaciones de variables escalares como la edad ya que si se usara la tabla sería inmensa e ilegible. Si queremos usar la edad como variable en una tabla de contingencia debemos recodificarla por rangos (p.ej. 18-35 años, 36-64 años, más de 64 años). Al recodificar una variable escalar como la edad por rangos, deja de ser escalar y pasa a ser ordinal, y por tanto sí se puede incluir un análisis de tablas de contingencia. Ejemplo:

Tabla 9. Tabla de contingencia: interés en la política según grupos de edad

		Grupos de edad			Total
		< 35	35-65	> 65	
Interés en la política	Muy interesado	53 10,0%	225 21,9%	148 30,5%	426 20,8%
	Algo interesado	218 41,1%	445 43,3%	188 38,7%	851 41,6%
	No muy interesado	160 30,2%	278 27,0%	130 26,7%	568 27,8%
	Nada Interesado	99 18,7%	80 7,8%	20 4,1%	199 9,7%
	Total	530 100,0%	1028 100,0%	486 100,0%	2044 100,0%

Fuente: elaboración propia a partir de Encuesta Mundial de Valores

<sup>3</sup> También se pueden usar las tablas de contingencia para analizar tres variables, pero son poco empleadas.



## **Qué variable en filas y qué variable en columnas**

Debido a que se estudia una variable en función de otra, el investigador ha de distinguir entre la variable dependiente (o a explicar) y la variable independiente (o explicativa). Esta distinción entre variable independiente y dependiente es importante porque la variable independiente se sitúa en columnas, y la variable dependiente en filas. El investigador es el que decide cuál variable es independiente (o explicativa) y cuál dependiente (o a explicar). Si nuestra hipótesis es causal o explicativa, la variable independiente (o explicativa) se coloca en columnas, y la variable dependiente (o a explicar) en filas. Si nuestra hipótesis es asociativa (y no causal o explicativa), el investigador debe decidir qué variable se coloca en filas y cuál en columnas. Las tablas de contingencia permiten comprobar hipótesis asociativas y causales (o explicativas). Aunque para afirmar causalidad deberían llevarse a cabo análisis más avanzados como tablas de contingencia de control (con 3 variables) o regresiones.

Las tablas de contingencia están compuestas de las categorías de respuesta de la variable en filas, de las categorías de respuesta de la variable en columnas, y el valor de las casillas. El valor de las casillas corresponde a la cantidad de casos que se adhieren a las respectivas categorías, y el porcentaje que representan estos casos.

## **Cómo se comprueba en una tabla de contingencia si dos variables están relacionadas**

Dos variables están relacionadas (asociadas) si la distribución de frecuencias de una variable (la que se coloca en filas) es diferente según las categorías de respuesta de la otra variable (la que se coloca en columnas). Es decir, si los resultados de una variable son diferentes en las categorías de respuesta de la otra variable es que las dos variables están relacionadas. En cambio, si los resultados de una variable son similares en las categorías de respuesta de la otra variable es que las variables son independientes (no están asociadas).

Para poder llevar a cabo esta comprobación es necesario que se calculen los porcentajes por columnas. Los porcentajes por columnas se obtienen de dividir el valor de una celda por el valor total de la columna. Por ejemplo, en la anterior tabla de contingencia donde se cruzan los datos de “interés en la política” y “grupos de edad”, hay 53 personas son menores de 35 años y están muy interesados en la política. Para calcular qué porcentaje representan, se divide 53 entre 530, que son el total de personas menores de 35 años. Así se obtiene el 10%, que representa el total de menos de 35 años que están muy interesados en la política. En la casilla de abajo, el 41,1% es el total de menores de 35 años que están algo interesados en la política, y se ha obtenido de dividir 218 (personas menores de 35 años algo interesadas en la política) entre 530 (total personas menores de 35 años). Este cálculo lo ejecutan los diversos programas informáticos, aunque siempre es el investigador el que debe indicar qué porcentaje se calcula y ser consciente de los datos que está interpretando.

Recomendación: las tablas de contingencia se leen fila por fila, y de derecha a izquierda. Esta es la mejor manera de poder averiguar si los valores de la variable en filas, se repiten por igual o de manera diferente en las categorías de respuesta de la variable en columnas. Si los valores (porcentajes) son muy diferentes es que la asociación entre las variables es fuerte. Si los valores son un poco diferentes es que la asociación entre las variables es débil. Y si los valores son muy similares o iguales es que no hay asociación entre variables. Se recomienda hacer una lectura fila por fila, y posteriormente una interpretación global de la asociación entre las dos variables.

## Ejemplo en 3 pasos:

### 1. Hipótesis de partida y selección de las dos variables

Quiero analizar si la creencia en Dios explica el interés en la política. Mi hipótesis de partida es que las personas creyentes en Dios tienen más interés en la política. Por tanto, tengo dos variables a relacionar: "interés en la política" y "creencia en Dios".

- La variable "interés en la política" fue preguntada en cuatro categorías de respuesta: 1-muy interesado, 2-algo interesado, 3-poco interesado, 4-nada interesado.
- La variable "creencia en Dios" tiene dos categorías: 1-sí creo, 2-no creo.

Voy a probar esta hipótesis usando la Encuesta Mundial de Valores realizada en Alemania en 2013. Antes de realizar la tabla de contingencia, estas son las tablas de frecuencias de cada variable por separado, sin haberlas cruzado o relacionado entre sí todavía.

Tabla 10. Tabla de frecuencias: Interés en la política

	Frecuencia	Porcentaje
Muy interesado	426	20,8%
Algo interesado	852	41,6%
No muy interesado	568	27,8%
Nada interesado	199	9,7%
Total	2045	100,0%

Tabla 11. Tabla de frecuencias: Creencia en Dios

	Frecuencia	Porcentaje
Sí creo	1286	65,1%
No creo	690	34,9%
Total	1976	100,0%

### 2. Construcción de la tabla de contingencia

Según mi hipótesis inicial es la creencia en Dios lo que explica el interés en la política, por tanto, la variable "creencia en Dios" será la variable independiente (o explicativa) y la variable "interés en la política" será la variable dependiente (o a explicar). La variable creencia en Dios irá en columnas y la variable interés en la política en filas.

Muy importante, al relacionar las dos variables en una tabla de contingencia, se calculan los porcentajes por columnas. ¿Cómo se calculan estos porcentajes por columnas? Normalmente estos cálculos los hace el programa informático con el que estemos

trabajando. Se detalla nuevamente cómo se calculan los porcentajes por columnas para que se pueda entender mejor el proceso y el resultado.

- Al construir una tabla de contingencia, en cada celda se coloca el total de casos que cumplen las categorías donde se cruzan. Por ejemplo, en la primera casilla, 283 es el número de personas que manifestaron que “sí creen en Dios” y además que dijeron tener “mucho interés en la política”.
- Para calcular los porcentajes por columnas, se divide el número de casos de cada casilla entre el total de casos de la columna. Por ejemplo, se divide el número de personas que “sí creen en Dios” y tienen “mucho interés en la política” entre el total de personas que “sí creen en Dios”, y se multiplica por 100 para expresarlo en porcentaje. En nuestro ejemplo:  $(283 / 1286) * 100 = 22,0\%$
- Seguidamente se calcula el porcentaje de todas las otras casillas. Por ejemplo, el número de personas que “no creen en Dios” y tienen “mucho interés en la política” se divide entre el total de personas que “no creen en Dios”. En nuestro ejemplo:  $(122 / 691) * 100 = 17,7\%$ . Y así con todas las casillas de la tabla.

Tras calcular los porcentajes por columnas, el resultado es la siguiente tabla de contingencia.

Tabla 12. Tabla de contingencia: Interés en la política según creencia en Dios

		Creencia en Dios		Total
		Sí creo	No creo	
Interés en la política	Muy interesado	283 22,0%	122 17,7%	405 20,5%
	Algo interesado	528 41,1%	306 44,3%	834 42,2%
	No muy interesado	356 27,7%	190 27,5%	546 27,6%
	Nada Interesado	119 9,3%	73 10,6%	192 9,7%
	Total	1286 100,0%	691 100,0%	1997 100,0%

Fuente: elaboración propia a partir de Encuesta Mundial de Valores

Una vez que la tabla de contingencia que relaciona dos variables está expresada en porcentajes ya podemos pasar a su lectura e interpretación. Si no están calculados los porcentajes, NO se puede leer la tabla de forma correcta ya que el número de casos no es igual en cada columna. Se deben calcular y presentar los porcentajes por columnas para analizar una tabla de contingencia.

#### 4. Lectura e interpretación de la tabla de contingencia

Recordemos que la tabla de contingencia se lee fila por fila, y de derecha a izquierda. Siguiendo con el ejemplo anterior, presentamos la manera correcta de describir esta tabla e interpretar los resultados.

El 20,5% de los encuestados en Alemania están muy interesados en la política. Este porcentaje es mayor en las personas que sí creen en Dios (22% de los que sí creen en Dios están muy interesados en la política, frente al 17,7 de los que no creen en Dios). Dijeron estar algo interesados en la política, el 42,2% de los encuestados, y este porcentaje es mayor para los que no creen en Dios. El 27,6 de los encuestados manifestaron estar poco interesados en la política, siendo este porcentaje casi igual para creyentes y no-creyentes en Dios (27,7 % y 27,5% respectivamente). Finalmente el 9,7% de los encuestados afirmaron no estar nada interesados en la política, siendo este porcentaje levemente mayor en los no-creyentes en Dios (9,3% de las personas que sí creen en Dios no están nada interesadas en política, frente al 10,6% de las que no creen en Dios). Por tanto, la creencia en Dios sí explica el interés en la política, aunque es una relación débil ya que las diferencias entre creyentes y no-creyentes en Dios, solo son medianamente considerables en la categoría de los muy interesados. En el resto de categorías del interés en la política, las diferencias entre los que sí creen en Dios y no creen en Dios, son pequeñas o inexistentes. Estos resultados nos ayudan a entender que el interés por la política no depende tanto de cuestiones de fe religiosa, y que el debate religioso no tiene mucha influencia sobre la movilización política. Futuros análisis deberían profundizar en otros aspectos para entender el interés en la política, tal vez el nivel educativo o el nivel de ingresos.

No hay una medida estándar para señalar si la diferencia entre porcentajes es alta o baja ya que las diferencias dependen del tamaño de la muestra y el número de categorías. Por ello, existen una serie de estadísticos denominados de contraste que sintetizan en un solo número si las dos variables están asociadas. Los estadísticos de contraste comprueban si las diferencias observadas para cada par de categorías de respuesta (casillas) difieren de las esperadas en caso de que las variables fueran independientes. El principal estadístico de contraste es chi-cuadrado de Pearson. Cuando la significación de chi-cuadrado es menor de 0,05 es que las dos variables están asociadas.

#### 5. Conclusiones:

- a. La tabla de contingencia es una técnica de análisis bivariado ya que relaciona dos variables y trata de averiguar si una variable explica la otra.
- b. Las tablas de contingencia permiten relacionar variables nominales y ordinales, pero no variables escalares (a no ser que las recodifiquemos por rangos).
- c. Hay que distinguir entre la variable que quiero explicar (variable dependiente) que irá en filas, y la variable explicativa (o independiente) que irá en columnas.
- d. Se debe calcular el porcentaje por columna para poder leer la tabla de contingencia.
- e. La tabla se lee fila por fila y de derecha a izquierda.
- f. Lo importante es averiguar si los porcentajes de la variable a explicar (la que va en filas) se diferencian mucho, poco o nada entre las categorías de la variable explicativa (la que va en columnas). Si hay altas diferencias de porcentajes las dos variables están relacionadas (asociadas), una variable explica la otra. Si no hay diferencias de porcentajes es que no hay relación entre las variables. Y si la

diferencia es pequeña u ocurre solo en algunas categorías es que la relación explicativa entre las variables es débil.

### Bibliografía sugerida

Sánchez Ramos, M. Á. (2005). Uso metodológico de las tablas de contingencia en la ciencia política. *Espacios Públicos*, 8(16), 60-84.

### Ejercicio

En la siguiente tabla de contingencia se presentan los resultados de cruzar las variables “Confianza en las grandes empresas” y “Género” basados en la encuesta Latinobarómetro realizada en Perú en 2012. Escriba un comentario e interpretación de la tabla donde se incluya:

- si hay asociación o no entre las variables,
- si existe asociación, señale si es fuerte, moderada o débil
- y qué tendencia o dirección tiene la asociación

Tabla 13. Tabla de contingencia: Confianza en las grandes empresas según género

		Género		Total
		Hombre	Mujer	
Confianza en las grandes empresas	Mucha confianza	57 9,6%	54 9,3%	111 9,5%
	Algo de confianza	176 29,7%	173 29,9%	350 29,8%
	No mucha confianza	214 36,1%	206 35,5%	420 35,8%
	Nada de confianza	145 24,5%	147 25,3%	292 24,9%
	Total	593 100,0%	580 100,0%	1173 100,0%

Solución: No existe asociación entre confianza en las grandes empresas y género. El 9,5% de los encuestados tienen mucha confianza en las grandes empresas, este porcentaje es casi idéntico para hombre y mujeres. El 29,8% del total de encuestados dicen tener algo de confianza en las empresas, y esta pauta es igual en hombres y mujeres. También es similar el porcentaje de respuestas entre los hombres y las mujeres que dicen tener no mucha confianza en las grandes empresas, y los que dicen no tener nada de confianza. En consecuencia, la confianza respecto a las grandes empresas es muy similar en hombres y mujeres, y por tanto no existe asociación significativa entre confianza en las grandes empresas y género. Otras variables explicativas que sí pueden estar asociadas a la confianza en las grandes empresas son el nivel de ingresos y el nivel educativo.

### 13. Correlaciones bivariadas

La correlación bivariada es una técnica estadística destinada a averiguar:

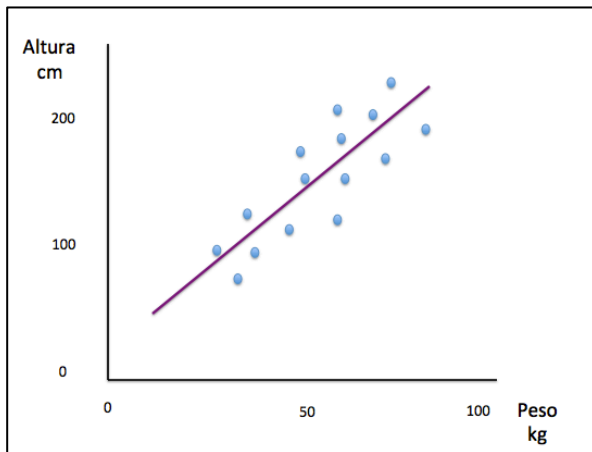
- a) si dos variables tienen relación entre sí
- b) si la relación es fuerte, moderada o débil y
- c) qué dirección tiene la relación

Las coincidencias muchas veces esconden asociaciones entre fenómenos. La correlación es la técnica más usada para medir asociación lineal en todas las ciencias.

El análisis de correlaciones indica asociación o relación entre dos variables, no implica causalidad.<sup>4</sup>

La correlación está basada en la asociación lineal, es decir, que cuando los valores de una variable aumentan los valores de la otra variable pueden aumentar o disminuir proporcionalmente. Por ejemplo, la altura y el peso corporal tienen una relación lineal positiva, a medida que aumenta la altura aumenta el peso. Si realizamos un gráfico de puntos con ambas variables, la nube de puntos se asemejará a una diagonal, lo cual indica que hay correlación entre esas dos variables.

Figura 6. Correlación entre altura y peso



Fuente: elaboración propia

Existen 2 grandes tipos de correlaciones: correlación de Pearson y correlación de Spearman. Ambas están basadas en la misma información, aunque usan fórmulas diferentes. La correlación de Pearson es más adecuada cuando las variables siguen la curva normal. La correlación de Spearman es más conveniente usarla cuando las variables no siguen la curva normal. Por lo general, no suele haber muchas diferencias entre los resultados, aunque pueden variar sobre todo cuando se trabaja con muestras pequeñas.

---

<sup>4</sup> Se pueden realizar análisis de correlaciones con tres variables, son denominadas correlaciones parciales, las cuales sí permiten indicar cierta causalidad. Aunque para demostrar causalidad se suelen usar análisis multivariados.

Se usa la correlación en análisis estadístico de datos cuando trabajamos con variables ordinales o escalares. Las variables ordinales y escalares son aquellas que sus categorías tienen un orden interno. Si incluimos una variable nominal debemos recodificarla a variable dummy. Una variable dummy es aquella de solo dos categorías o valores, 1 y 0, la categoría o valor 1 indica presencia del fenómeno, y la categoría o valor 0 indica ausencia del fenómeno.

### Cómo analizar la correlación bivariada en 2 pasos

La gran ventaja de la correlación es que toda la información de existencia de relación, fortaleza y dirección, aparece sintetizada en un coeficiente de correlación ( $r$ ) y un nivel de significación ( $sig.$ ).

1. El nivel de significación indica si existe o no relación entre dos variables. El nivel de significación más usado es 0,05, lo cual hace referencia al 95% de nivel de confianza, que es la probabilidad de que el resultado no se deba al azar. Cuando la significación obtenida de correlacionar dos variables es menor de 0,05 sí existe correlación significativa entre esas dos variables. Si existe correlación significativa debemos pasar al paso 2.

2. El coeficiente de correlación ( $r$ ) indica que tan fuerte o débil es una correlación. Este coeficiente puede oscilar entre -1 y +1. Cuanto más se aleja de 0, más fuerte es la relación entre las dos variables. Cuanto más cerca de 0 es que la relación entre las variables es débil. Si el coeficiente de correlación es muy próximo a 0 es que ambas variables no están correlacionadas. Por otro lado, el signo (positivo o negativo) del coeficiente de correlación indica la dirección de la relación.

Varios ejemplos para entenderlo mejor:

#### Ejemplo 1:

La muestra ( $N$ ) son: 2249 encuestados en Colombia (World Values Survey 2005)

Analizamos la relación entre “Ideología” e “Importancia de Dios en la vida”

- Ideología es una escala de 1 a 10, donde 1 es extrema izquierda y 10 es extrema derecha.
- Importancia de Dios en la vida es una escala donde 1 es nada importante y 10 muy importante.

		Ideología
Importancia de Dios en la vida	Correlación de Pearson ( $r$ )	0,124
	Significación	0,000
	$N$	2249

Existe correlación significativa entre Ideología e Importancia de Dios en la vida ya que la significación es 0,000 y por tanto menor de 0,05, por lo que este resultado no se debe al azar. La correlación de Pearson ( $r = 0,124$ ) señala que se trata de una relación débil al estar próxima a 0. El signo positivo de esta correlación señala que a cuanto más de derecha son las personas en Colombia, más importancia le dan a Dios en la vida. Y lo mismo si

lo leemos de forma inversa, cuanto menos importancia de dan a Dios en la vida, más de izquierda son las personas colombianas.

**Advertencia:** hay que tener muy en cuenta cómo están ordenadas las categorías de las variables, ya que la lectura de la dirección de la relación está basada en el orden de las categorías. Es decir, el signo positivo de una correlación nos señala que los resultados se asemejan a una diagonal hacia arriba, pero es el investigador el que debe comprobar cómo se interpretan los resultados. Por ejemplo, la importancia de Dios en la vida fue preguntado en una escala de 1 a 10 donde 1 es nada importante y 10 es muy importante, y la ideología en una escala del 1 al 10 donde 1 es extrema izquierda y 10 es extrema derecha. El valor de la correlación de Pearson ( $r = 0,124$ ) al ser positivo indica que, a más importancia de Dios en la vida, más de derechas son las personas, o lo que es lo mismo, a menor importancia de Dios en la vida, más tendencia a ser de izquierdas. Pero si la variable importancia de Dios en la vida hubiera sido preguntado en una escala de 1 a 10, pero donde el 1 es muy importante y el 10 es nada importante (al revés de como se hizo originariamente), el valor de la correlación de Pearson entre importancia de Dios en la vida e ideología hubiera sido ( $r = -0,124$ ), es decir, mismo valor, pero con signo negativo. Esto indica que al aumentar los valores de la variable importancia de Dios en la vida, descenden los valores de la variable ideología. Como ideología está preguntado de izquierda a derecha, e importancia de Dios en la vida de muy importante a nada importante, la lectura sería: a más importancia de Dios en la vida, más de derechas son las personas, o bien menor importancia de Dios en la vida, más de izquierdas son las personas. Es decir, el resultado es el mismo, pero hemos de estar muy atentos al orden de las categorías de las variables para leer correctamente la dirección de una correlación significativa.

### Ejemplo 2:

La muestra (N): 3017 personas en Colombia (World Values Survey 2005)

Analizamos la correlación entre “Edad” e “Interés en la política”

- La edad es una variable escalar
- Interés en la política es una variable ordinal donde las categorías son: 1-mucho interés, 2-bastante interés, 3-poco interés, 4-nada de interés

		Interés en la política
Edad	Correlación de Pearson (r)	0,013
	Significación	0,467
	N	3017

No hay correlación significativa entre edad e interés en la política ya que la significación es mayor de 0,05 (Sig. = 0,467). A medida que aumenta la edad no crece o decrece el interés en la política. Por tanto, deberíamos buscar otras variables si queremos



comprender con qué se relaciona el interés en la política, ya que la edad no correlaciona con el interés en la política.

El uso de la correlación es útil para caracterizar y extraer perfiles. Por ejemplo, permitiría identificar quiénes son los que más le dan importancia a Dios en la vida. De momento ya sabemos que los que tienen tendencia política más a la derecha. ¿Qué otras variables correlacionan con ésta? Las variables que correlacionen con la de nuestro interés serán las que nos permita identificar perfiles. Además, las correlaciones permiten analizar relaciones entre fenómenos (o variables). Por ejemplo, ¿hay relación entre la inversión en educación y la reducción de crímenes? ¿A más turismo extranjero más reducción de la pobreza? ¿A más número de becas concedidas menor satisfacción con el gobierno? Investigar es descubrir relaciones entre fenómenos, y las correlaciones son imprescindibles para ello.

### **Bibliografía sugerida**

Llopis Pérez, J. (2012, noviembre 30). Tema 5: Correlación. Recuperado a partir de <https://estadisticaorquestainstrumento.wordpress.com/2012/11/30/tema-4-correlacion/>

### **Ejercicio**

El siguiente cuadro presenta los resultados de calcular las correlaciones entre varias variables. A pesar que se presenten todas las variables en un cuadro, la correlación es una técnica bivariada donde se comprueba la asociación entre solo dos variables. Las variables analizadas y sus respectivas categorías de respuesta son:

- Nivel de felicidad: 1. Nada feliz, 2. Algo feliz, 3. Bastante feliz, 4. Muy feliz
- Importancia de Dios en la vida: 1. Nada importante, 2, 3, 4, 5, 6, 7, 8, 9, 10. Muy importante
- Interés en la política: 1. Muy interesado, 2. Algo interesado, 3. Poco interesado, 4. Nada interesado
- Nivel educativo: 1. Sin estudios, 2. Primarios, 3. Secundarios, 4. Universitarios

Revise la tabla y comente todos resultados. Debe incluir para cada análisis de correlación respuesta a estas preguntas:

- a) ¿Existe asociación (o correlación) significativa entre las 2 variables?
- b) Si existe asociación, ¿es una relación fuerte, moderada o débil?
- c) Si existe asociación, ¿qué dirección tiene esta correlación?

Tabla 14. Tabla de correlaciones de Pearson

		Nivel de felicidad	Importancia de Dios en la vida	Nivel educativo	Número de hijos
Importancia de Dios en la vida	Correlación de Pearson	0,035			
	Significación	0,230			
	N	1199			
Nivel educativo	Correlación de Pearson	0,168	-0,116		
	Significación	0,000	0,000		
	N	1202	1207		
Número de hijos	Correlación de Pearson	-0,085	0,104	-0,358	
	Significación	0,003	0,000	0,000	
	N	1197	1202	1205	
Interés en la política	Correlación de Pearson	0,059	-0,062	0,213	-0,114
	Significación	0,041	0,033	0,000	0,000
	N	1193	1197	1200	1195

- No existe correlación significativa entre nivel de felicidad e importancia de Dios en la vida ya que el nivel de significación es mayor de 0,05.

- Si existe correlación significativa entre nivel de felicidad y nivel educativo ya que el nivel de significación es menor de 0,05. La correlación de Pearson es 0,168. Comparando este valor de nivel de felicidad y nivel educativa es moderada (ni muy fuerte ni tampoco muy débil). El signo positivo de la correlación de Pearson indica que a más nivel educativo más nivel de felicidad, o lo que es lo mismo, a menor nivel educativo, menor nivel de felicidad.

- Si existe asociación (o correlación) significativa entre nivel de felicidad y número de hijos ya que la significación es menor de 0,05. El valor de la correlación de Pearson es muy próximo a 0 y comparándolo con otros valores de correlación de esta misma tabla nos indica que es una relación débil. El signo negativo nos indica que a mayor número de hijos, menor nivel de felicidad, o lo que es lo mismo, a menos o ningún hijo, más nivel de felicidad, aunque cabe tener en cuenta que es una relación débil.

Solución:

## 14. Qué es y cómo analizar ANOVA de un factor

ANOVA de un factor (también llamada ANOVA unifactorial o one-way ANOVA en inglés) es una técnica estadística que señala si dos variables (una independiente y otra dependiente) están relacionadas en base a si las medias de la variable dependiente varían en las categorías o grupos de la variable independiente. Es decir, señala si las medias entre dos o más grupos son similares o diferentes.

ANOVA son las siglas de Analysis of Variance. Hay varios subtipos de ANOVA. Nos centramos en ANOVA de un factor ya que es un tipo de análisis bivariado ya que comprueba si dos variables están asociadas. Se le denomina ANOVA de un factor porque a la variable independiente se le conoce como factor.

### 1. ¿Cuándo usar ANOVA de un factor?

Usamos ANOVA de un factor cuando queremos saber si las medias de una variable son diferentes entre los niveles o grupos de otra variable. Por ejemplo, si queremos comparar el número promedio de hijos entre los grupos o niveles de clase social: clase baja, clase trabajadora, clase media-baja, clase media-alta y clase alta. Es decir, vamos a comprobar mediante ANOVA si la variable “número de hijos” está relacionada con la variable “clase social”. Concretamente, se analizará si la media del número de hijos varía según el nivel de clase social a la que pertenece la persona.

#### Condiciones:

- En ANOVA de un factor solo se relacionan dos variables: una variable dependiente (o a explicar) y una variable independiente (o explicativa, que en esta técnica se suele llamar factor)
- La variable dependiente es cuantitativa (escalar, u ordinal de más de 5 categorías) y la variable independiente es categórica (nominal u ordinal).
- Se pide que las variables sigan la distribución normal, aunque como siempre esto es difícil de cumplir en investigaciones sociales.
- También que las varianzas (es decir, las desviaciones estándar al cuadrado) de cada grupo de la variable independiente sean similares (fenómeno que se conoce como homocedasticidad). Aunque esto es lo ideal, en la realidad cuesta de cumplir, e igualmente se puede aplicar ANOVA, aunque los resultados deben ser interpretados con mayor precaución.

### 2. ¿En qué se basa ANOVA de un factor?

ANOVA de un factor compara las medias de la variable dependiente entre los grupos o categorías de la variable independiente. Por ejemplo, comparamos las medias de la variable “Número de hijos” según los grupos o categorías de la variable “Clase social”.

Si las medias de la variable dependiente son iguales en cada grupo o categoría de la variable independiente, los grupos no difieren en la variable dependiente, y por tanto no hay relación entre las variables. En cambio, y siguiendo con el ejemplo, si las medias del número de hijos son diferentes entre los niveles de la clase social es que las variables están relacionadas.

### 3. ¿Qué estadísticos se calculan en ANOVA?

Al aplicar ANOVA de un factor se calcula un estadístico o test denominado F y su significación. El estadístico F o F-test (se llama F en honor al estadístico Ronald Fisher) se obtiene al estimar la variación de las medias entre los grupos o categorías de la variable independiente y dividirla por la estimación de la variación de las medias dentro de los grupos. El cálculo del estadístico F es algo complejo de entender, pero lo que hace es dividir la variación entre los grupos por la variación dentro de los grupos. Si las medias entre los grupos varían mucho y la media dentro de un grupo varía poco, es decir, los grupos son heterogéneos entre ellos y similares internamente, el valor de F será más alto, y por tanto, las variables estarán relacionadas. En conclusión, cuanto más difieren las medias de la variable dependiente entre los grupos o categorías de la variable independiente, más alto será el valor de F. Si hacemos varios análisis de ANOVA de un factor, aquel con un valor de F más alto indica que hay más diferencias, y por tanto una relación más fuerte entre las variables. Para decir que el valor de F es alto o medio o bajo, debemos compararlo con otro análisis realizado con los mismos datos, o datos similares.

### 4. ¿Cómo se interpreta el test de F y la significación?

La significación de F se interpreta como la probabilidad de que este valor de F se deba al azar. Siguiendo un nivel de confianza del 95%, el más utilizado en ciencias sociales, cuando la significación de F sea menor de 0,05 es que las dos variables están relacionadas.

Hemos de analizar e interpretar al aplicar ANOVA de un factor:

- **Significación:** si es menor de 0,05 es que las dos variables están relacionadas y por tanto que hay diferencias significativas entre los grupos.
- **Valor de F:** cuanto más alto sea F, más están relacionadas las variables, lo que significa que las medias de la variable dependiente difieren o varían mucho entre los grupos de la variable independiente.

### 5. Ejemplos de ANOVA de un factor

#### Ejemplo 1:

Quiero averiguar si el número de hijos varían según la clase social. Para ello comparo las medias de números de hijos entre los diversos niveles, grupos o categorías de clase social: clase baja, clase trabajadora, clase media baja, clase media-alta, y clase alta. Utilizo los datos de la Encuesta Mundial de Valores realizada entre 2010 y 2014 en 58 países del mundo.

Introduciré las dos variables que quiero analizar:

- “Número de hijos”: es una variable cuantitativa que va de 0 hasta 8 o más hijos. Esta será la variable dependiente o a explicar
- “Clase social”: variable ordinal. Es la variable independiente o factor. Las categorías o grupos de la clase social son: clase baja, clase trabajadora, clase media baja, clase media-alta, y clase alta.

El resultado es el siguiente:

En la tercera columna se observan las medias para cada grupo de clase social. Si nos fijamos en las medias del número de hijos en cada grupo de clase social, podemos observar que a medida que aumenta la clase social desciende la media del número de

hijos. Las personas de clase baja tienen de media 2,17 hijos, las de clase trabajadora 1,93 hijos, las de clase media-baja 1,82 hijos en promedio, las de clase media-alta tienen 1,74 hijos de media, y las de clase alta tienen de media 1,76 hijos.

Tabla 15. Descriptivos: Número de hijos según clase social

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mín	Máx
					Límite inferior	Límite superior		
Clase alta	1938	1,76	1,798	0,041	1,68	1,84	0	8
Clase media-alta	17889	1,74	1,735	0,013	1,71	1,76	0	8
Clase media-baja	31389	1,82	1,729	0,01	1,8	1,84	0	8
Clase trabajadora	24333	1,93	1,76	0,011	1,91	1,96	0	8
Clase baja	10760	2,17	2,014	0,019	2,14	2,21	0	8
Total	86309	1,88	1,783	0,006	1,87	1,89	0	8

Tabla 16. ANOVA de un factor

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	1526,009	4	381,502	120,622	0
Intra-grupos	272961,208	86304	3,163		
Total	274487,217	86308			

El valor de F es 120,622 y la significación es 0,000. Al ser la significación menor de 0,05 es que las diferencias de la media de hijos entre los grupos de la clase social son significativas. Aunque aparentemente podemos pensar que las diferencias no son exageradas, la decisión de si las diferencias son significativas no depende de nuestro criterio, sino de la significación de F. Este es el objetivo de aplicar ANOVA de un factor: valorar estadísticamente si las diferencias de medias son significativas o no.

### Ejemplo 2:

Relacionamos la variable "Número de hijos" con la variable "¿Es una persona religiosa?". Se les preguntó a las personas si eran religiosos y se les dio 3 opciones de respuesta:

1. soy es una persona religiosa
2. no soy una persona religiosa
3. soy ateo

Comparamos la media de hijos por los grupos de la variable "¿Es una persona religiosa?". En la tabla de descriptivos nos fijamos en la columna de las medias, y observamos que

las personas religiosas tienen de media 2,06 hijos, las no religiosas tienen de media 1,56 hijos y los ateos 1,3 hijos en promedio.

Tabla 17. Descriptivos: Número de hijos según es una persona religiosa

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mín	Máx
					Límite inferior	Límite superior		
Soy una persona religiosa	57781	2,06	1,878	0,008	2,04	2,07	0	8
No soy una persona religiosa	22186	1,56	1,545	0,01	1,54	1,58	0	8
Soy ateo	4199	1,3	1,289	0,02	1,27	1,34	0	8
Total	84166	1,89	1,789	0,006	1,87	1,9	0	8

Tabla 18. ANOVA de un factor

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	5496,42	2	2748,21	876,775	0
Intra-grupos	263804,841	84163	3,134		
Total	269301,262	84165			

La significación de F es 0,000, al ser menor de 0,05 es que hay relación significativa entre las variables, y el valor de F es 876,775.

Podemos comparar el valor de F de este ejemplo con el del ejemplo anterior ya que están basados en los mismos datos:

- "Número de hijos" según "Clase Social"  $F = 120,622$
- "Número de hijos" según "¿Es una persona religiosa?"  $F = 876,775$

El valor de F es más grande en la relación de "Número de hijos" y "¿Es una persona religiosa?" porque las diferencias de medias del número de hijos son más grandes que las diferencias de medias según la clase social. Cuanto más alto sea el valor de F, más diferencias de medias habrá entre los grupos o categorías de la variable independiente (o factor) y por tanto más fuerte es la relación entre las variables.

## 6. Conclusión

ANOVA de un factor se utiliza muchísimo en las ciencias sociales. Es muy popular en psicología y en análisis comparativo y experimental para poder saber si las diferencias de un grupo respecto a otro son significativas y qué fortaleza tienen.

A la hora de presentar ANOVA en un informe, artículo de investigación o tesis, se debe presentar la tabla de las medias y desviaciones típicas, y seguidamente el estadístico F y su significación.

Sin ANOVA de un factor, las diferencias entre un grupo y otro quedarían a juicio subjetivo del observador, y donde una persona ve diferencias otra quizás no las verías. Es mejor usar la estadística para saber si hay similitud o diferencia entre los grupos, y ANOVA es excelente para este propósito.

### Ejercicio

En la Encuesta Mundial de Valores se preguntó a las personas en qué nivel toleran la prostitución en una escala del 1 al 10, donde 1 es nunca lo justificaría y 10 es siempre lo justificaría. Se analizan las diferencias de opinión respecto a la “Tolerancia con la prostitución” según la “clase social” de los encuestados, y posteriormente según el “género” de los encuestados. Se aplicó ANOVA de un factor, y estos son los resultados.

Comente cada análisis de ANOVA y responda después qué variable independiente (o factor) está más relacionada con la “Tolerancia con la prostitución”.

Tabla 19. Descriptivos: Tolerancia con la prostitución según clase social

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Clase alta	1414	2,91	2,672	0,071	2,77	3,05	1	10
Clase media-alta	12769	2,92	2,565	0,023	2,88	2,97	1	10
Clase media-baja	22310	2,81	2,493	0,017	2,77	2,84	1	10
Clase trabajadora	17913	2,64	2,429	0,018	2,6	2,67	1	10
Clase baja	8268	2,56	2,428	0,027	2,5	2,61	1	10
Total	62674	2,75	2,489	0,01	2,73	2,77	1	10

Tabla 20. ANOVA de un factor: Tolerancia con la prostitución según clase social

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	1023,008	4	255,752	41,39	0
Intra-grupos	387240,042	62669	6,179		
Total	388263,05	62673			

Tabla 21. Descriptivos: Tolerancia con la prostitución según género

	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Hombre	30613	2,88	2,578	0,015	2,85	2,91	1	10
Mujer	33293	2,65	2,407	0,013	2,62	2,67	1	10
Total	63906	2,76	2,493	0,01	2,74	2,78	1	10

Tabla 22. ANOVA de un factor: Tolerancia con la prostitución según género

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	860,138	1	860,138	138,68	0
Intra-grupos	396352,437	63904	6,202		
Total	397212,575	63905			

- Las respuestas del total de personas encuestadas en cuanto a la tolerancia con la prostitución en una escala del 1 al 10 donde 1 es nunca y 10 es siempre, se sitúan en promedio en 2,75 (desviación estándar 2,489). Existen diferencias según la clase social. Las clases alta y media-alta tienen más tolerancia hacia la prostitución ya que el promedio es mayor que el de las clases baja y trabajadora. Estas diferencias de medias son significativas según el análisis de ANOVA de un factor ( $F = 41,39$ ;  $Sig. = 0,000$ ).
- Si analizamos las diferencias en la tolerancia con la prostitución según el género de los encuestados, observamos que los hombres tienden a ser más tolerantes que las mujeres en promedio, 2,88 los hombres y 2,65 las mujeres. Estas diferencias son significativas según el análisis de ANOVA de un factor ( $F = 138,68$ ;  $Sig. = 0,000$ ).
- El valor más alto de F indica que las diferencias respecto a la tolerancia con la prostitución según el género son mayores que según la clase social. Es decir, hay más variación de opiniones según el género que según la clase social de las personas. Aunque tanto género como clase social están relacionadas con la tolerancia con la prostitución, el género es más explicativo que la clase social para entender las diferencias de opiniones respecto a la prostitución.

Solución:



## 15. Presentación de los resultados de investigación

Toda investigación debe finalizar con la presentación de los resultados.

Casi todas las formas de presentación de los resultados deben incluir las siguientes secciones:

- Título – cuanto más sugerente y al mismo tiempo descriptivo mejor. Hay que tener el uso de palabras claves relacionadas con nuestro campo de estudio para que los buscadores puedan situar nuestro trabajo donde queremos. Por ejemplo, si analizamos la crisis financiera hay que usar este término en el título y el resumen y no similitudes como “gran depresión” o “malos tiempos”.
- Resumen o abstract – se debe exponer en un máximo de 250 palabras el objetivo de la investigación, los métodos, datos empleados y los resultados obtenidos. El resumen es una pieza clave en la investigación ya que es la primera impresión o carta de presentación de nuestro trabajo. Las personas decidimos si vamos a leer el informe o artículo de investigación en función de la calidad del resumen y del interés que nos despierte, por ello debe estar impecable. Además, el resumen se usa como mecanismo de filtro y selección de trabajos en una convocatoria para participar en congreso, conferencia o incluso concurso.
- Introducción – en esta sección se debe incluir la definición y delimitación del problema de investigación, y la pregunta de investigación u objetivo. Se deben responder estas cuestiones: qué voy a investigar, por qué investigar ese problema, para qué investigarlo, dónde voy a investigarlo, cuándo voy a investigarlo, cuál es el aporte de mi investigación o vacío que cubrirá, y qué queremos averiguar o saber (la pregunta de investigación). La introducción es la parte fundamental de una presentación de investigación ya es el momento en que el evaluador o lector tiene el mayor grado de atención, y suele juzgar un trabajo de investigación en función de la claridad y calidad de la introducción. En esta sección debe proveerse toda la información importante de una investigación y debe dejar claro para todo lector qué se quiere investigar.
- Estado del arte – lugar donde se exponen las investigaciones anteriores que analizaron o teorizaron sobre el problema de investigación. Dentro de esta sección puede incluirse un marco conceptual donde se definan los principales conceptos de nuestra investigación. Si se utiliza alguna teoría para analizar el problema de investigación, desarrollaremos en esta sección un marco teórico. Es recomendable sintetizar mucha toda esta información sobre investigaciones previas y teorías para que no parezca nuestro trabajo solo un sinfín de comentarios de otras investigaciones. El investigador debe ir dialogando con las teorías o trabajos anteriores, por ejemplo, comentando sus falencias o aspectos a mejorar. No debe ser solo un resumen de qué han dicho otros, sino un diálogo. Hemos de tener en cuenta que lo importante en la presentación de los resultados es nuestra investigación, y no las investigaciones que hicieron otros antes.
- Métodos y datos– en esta sección debe exponerse en detalle los instrumentos y datos que usamos para responder la pregunta de investigación o alcanzar el objetivo. Se debe justificar la elección del método empleado y las posibles limitaciones. Se debe explicar la selección de los casos o datos usados, justificando por qué se seleccionaron esos casos y esas variables para responder la pregunta de investigación.

- Análisis y resultados – espacio donde presentar los principales resultados obtenidos. Se deben detallar los análisis realizados mediante la presentación de tablas y/o gráficos de resultados. En esta sección no se suele coloca bibliografía o referencias, sino solo presentar los resultados que nos ayuden a responder la pregunta de investigación.
- Discusión – lugar para la interpretación de los resultados, donde el investigador debe señalar todo aquello que le sugieren los resultados obtenidos. Por ejemplo, relacionar los resultados con teorías o con otras investigaciones, o realizar otras lecturas de los resultados más allá de la pregunta de investigación u objetivo de la investigación.
- Conclusiones – sección para responder la pregunta de investigación y señalar futuras líneas de investigación, y si cabe posibles limitaciones del estudio.
- Referencias o bibliografía – se deben incluir todos los trabajos citados a lo largo del texto. Existen cientos de formas diferentes de presentar la bibliografía. Se debe usar aquel criterio que nos demande la revista o institución a quién dirigimos la investigación. Si no existe ningún criterio establecido, lo relevante es que todas las referencias sigan el mismo criterio. No puede citarse un artículo mediante el criterio APA y otro artículo mediante el criterio SAGE Harvard. Por ello, se recomienda el uso de algún gestor bibliográfico como Zotero o Endnote para facilitar la presentación de las referencias.

Las principales formas de presentación de los resultados son:

- Presentación de una comunicación en un congreso – útil para presentar resultados parciales o trabajos en progreso mediante una comunicación de aproximadamente 20 minutos. Debe enfocarse en los resultados principales y no tanto en el estado del arte o la metodología.
- Artículo de investigación – escrito de un máximo de 8.000 o 10.000 palabras (unas 20 páginas) donde se presenta la investigación en búsqueda de difusión en el sector académico.
- Informe de investigación – realizado cuando la investigación es encargada por alguna institución. Tiene una forma algo diversa al del artículo académico. En el informe de investigación se prioriza la presentación de los resultados y posibles recomendaciones tras el análisis realizado.
- Tesis – presentación de los resultados de una investigación para obtener un título académico. La extensión varía entre las 100 y 400 páginas según sea para un Máster o para Doctorado.

Se recomienda siempre revisar trabajos que hayan sido altamente citados o que hayan obtenido un alto reconocimiento para usarlos como modelos a seguir.

### **Ejercicio**

Busque cuáles son las revistas científicas con mayor índice de impacto en su campo y seleccione uno de los últimos artículos publicados. Revise minuciosamente ese artículo y compruebe cómo plantea el problema de investigación, cómo presenta la pregunta u objetivo de investigación, cómo estructura el estado del arte, cómo expone el método y datos empleados, cómo presenta los resultados y qué aspectos destaca en la discusión y conclusiones.

No olvide que a investigar se aprende investigando, pero también leyendo investigaciones. En el mundo de la investigación nunca se deja de aprender. La investigación es eterna.

## Referencias

Aprender LyX. Planteamiento del Problema, Definición - Partes del Planteamiento de un Problema. Disponible en: <https://www.youtube.com/watch?v=jnqONgntR6g>

Cea d'Ancona, M.A. (1999) Metodología Cuantitativa: Estrategias y Técnicas de Investigación Social. Capítulo 7. Madrid: Síntesis. Disponible en: [http://www.trabajosocial.unlp.edu.ar/uploads/docs/metodologia\\_cuantitativa\\_estrategias\\_y\\_tecnicas\\_de\\_investigacion\\_social\\_cea\\_d\\_ancona.pdf](http://www.trabajosocial.unlp.edu.ar/uploads/docs/metodologia_cuantitativa_estrategias_y_tecnicas_de_investigacion_social_cea_d_ancona.pdf)

Cea d'Ancona, M.A. (1999) Metodología Cuantitativa: Estrategias y Técnicas de Investigación Social. Capítulo 4. Madrid: Síntesis. Disponible en: <http://metodos-comunicacion.sociales.uba.ar/files/2014/04/Cea-D-Ancona-Cap-3-y-4.pdf>

Díaz de Rada, V. (2001) *Diseño y elaboración de cuestionarios para la investigación comercial*. Madrid: ESIC Editorial

Encuesta Mundial de Valores. Disponible en: <http://www.worldvaluessurvey.org/WVSContents.jsp>

Enfoque Cualitativo, Cuantitativo y Mixto. Disponible en: <https://www.youtube.com/watch?v=XpODjddaAqg>

Estadística para torpes. Disponible en: <https://unidadinvestigacionhvn.files.wordpress.com/2010/11/estadistica-para-torpes.pdf>

Ferrans, C.E. (1990) Development of a quality of life index for patients with cancer. *Oncology Nursing* 17: 15–21.

Gómez L.E, Verdugo, M.A., Arias, B. & Navas, P. (2008) Evaluación de la calidad de vida en personas mayores y con discapacidad: la Escala Fumat. *Psychosocial Intervention* 17(2): 189–199.

Latinobarómetro. Disponible en: <http://www.latinobarometro.org/latOnline.jsp>

Lawton, M. (2001) Quality of life in chronic illness. *Gerontology* 45: 181–183.

Llopis Pérez, J. (2012, noviembre 30). Tema 5: Correlación. Disponible en: <https://estadisticaorquestainstrumento.wordpress.com/2012/11/30/tema-4-correlacion/>

Mallou, J.V., Tobío, T.B., Carreira, A.G., Boubeta, A.R. & Fernández, X.G.V. (1998). Estimación de la respuesta de los «no sabe/no contesta» en los estudios de intención de voto. *REIS*, 83, 269-287. <https://doi.org/10.2307/40184128>

Reguant, M. & Martínez-Olmo, F. (2014). Operacionalización de conceptos / variables. Barcelona: Dipòsit Digital de la UB. Disponible en: <http://diposit.ub.edu/dspace/bitstream/2445/57883/1/Indicadores-Repositorio.pdf>

Salmerón Gómez, Romelio (2015) (Mini)Manual de PSPP, alternativa libre a SPSS. Disponible en: <https://softtcm.files.wordpress.com/2014/04/pspp.pdf>

Sánchez Ramos, M.Á. (2005). Uso metodológico de las tablas de contingencia en la ciencia política. *Espacios Públicos*, 8(16), 60-84.

Urzúa M.Á. & Caqueo-Urizar A. (2012) Calidad de vida: Una revisión teórica del concepto. *Terapia psicológica* 30(1): 61–71.

## **Sobre trAndeS**

**trAndeS** es un programa de postgrado estructurado con base en la Pontificia Universidad Católica del Perú (PUCP) que contribuye al desarrollo sostenible en la región andina a través de sus actividades de investigación y formación. Los socios del proyecto son Freie Universität Berlin y Pontificia Universidad Católica del Perú (PUCP).

**trAndeS** está financiado por el Servicio de Intercambio Académico Alemán (Deutscher Akademischer Austauschdienst, DAAD), con fondos del Ministerio Federal de Cooperación Económica y Desarrollo (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, BMZ).

El objetivo de **trAndeS** es crear y promover conocimientos que puedan contribuir al logro de la Agenda 2030 de Naciones Unidas con sus 17 Objetivos de Desarrollo Sostenible (ODS) en la región andina. En particular, el programa pretende identificar cómo las persistentes desigualdades sociales en la región presentan desafíos para alcanzar los ODS y cómo el progreso hacia estos objetivos pueda contribuir a la reducción de las desigualdades.

Más información en [www.programa-trandes.net](http://www.programa-trandes.net).

## Instituciones ejecutivas de trAndeS



### Berlín

Freie Universität Berlin  
Instituto de Estudios Latinoamericanos  
Boltzmannstr. 1  
14195 Berlin  
T: +49 30 838 53069  
contacto@programa-trandes.net



### Lima

Pontificia Universidad Católica del Perú  
Departamento de Ciencias Sociales  
Universitaria 1801  
Lima 32, Peru  
T: +51 1 626 2000 Ext. 5138  
trandes@pucp.edu.pe

# DAAD

Con el apoyo financiero del



Ministerio Federal de  
Cooperación Económica  
y Desarrollo