

## T. 1 – Introducción a la Estadística aplicada a la Psicología

1. La Estadística
2. El método científico y la Estadística
3. ¿Por qué la Estadística en el grado de Psicología?
4. Algunos conceptos básicos de Estadística
5. Metodologías de investigación y Estadística
6. Estadística descriptiva y estadística inferencial
  - 6.1. Población y muestra
  - 6.2. Parámetros y estadísticos

### 1. La Estadística

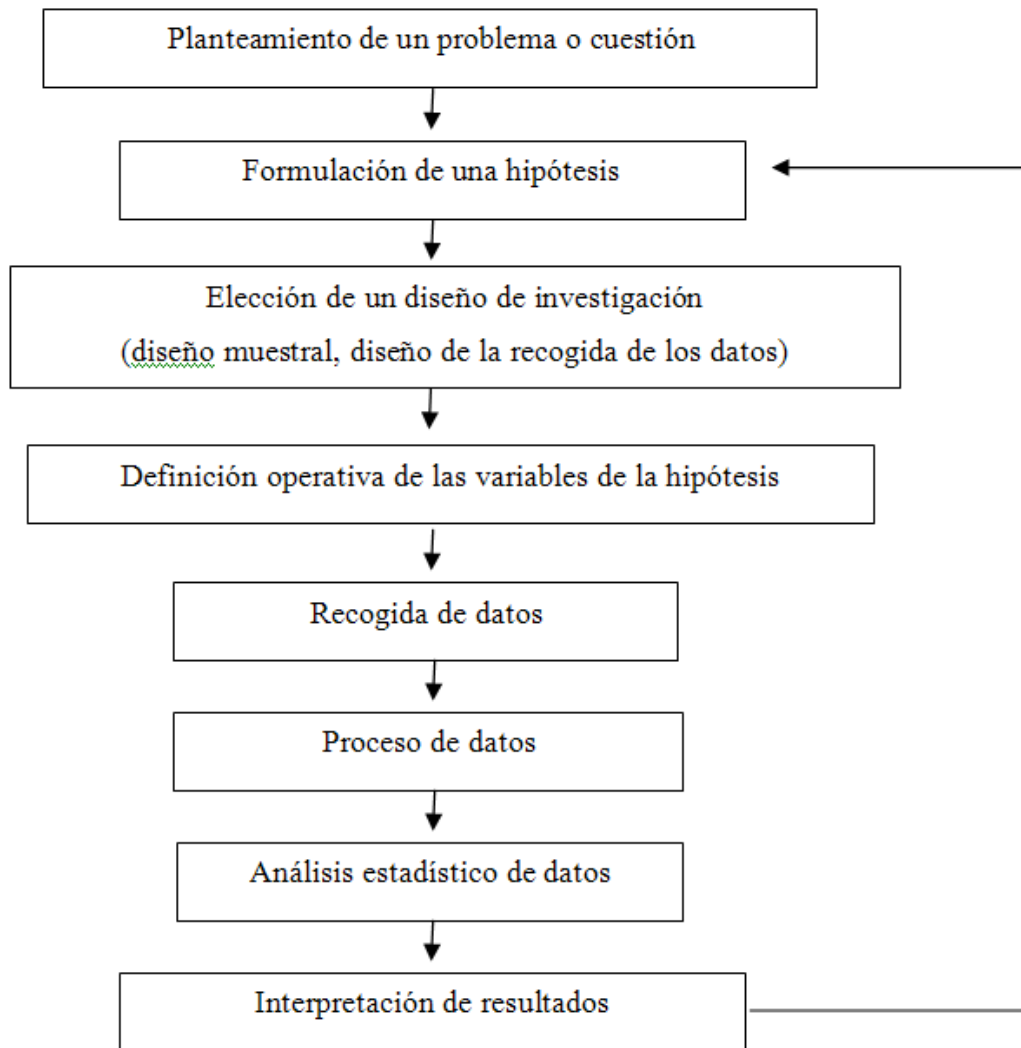
- Tiene su origen en el interés de los *Estados* por conocer los recursos con que contaban: número de habitantes, edad, tipo de trabajo que realizaban, condiciones de vida, propiedades, etc. En los imperios romano y egipcio ya aparecen manifestaciones de ese interés por procedimientos que permitieran obtener datos estadísticos (*de estado*). En cualquier caso, va a ser en el periodo de dominio napoleónico cuando se produzca el salto más sustancial en ese afán por disponer de información estadística. Posteriormente, el interés se ha desplazado a otros niveles de análisis más moleculares (ciudades, barrios, colegios, grupos concretos de personas...) o también, a veces, más molares (grupos de naciones, continentes, el mundo...). Por otra parte, el interés del análisis se ha extendido a todo tipo de variables más allá de las que típicamente cubre el censo, el heredero directo de aquel empeño remoto en que se puede encontrar el origen de la estadística.

- Aunque la estadística aparezca originalmente asociada al interés por conocer acerca de los habitantes de naciones, regiones u otros tipos de agrupaciones, pronto trascendió su aplicación a otro tipo de unidades no necesariamente humanas, pero en los que era también habitual la recogida de volúmenes amplios de datos de los que se quería extraer la información esencial –este es el caso, por ejemplo, de la investigación astronómica o, más reciente en el tiempo, la investigación acerca de productos de consumo.

- Ese interés por la estadística ha dado lugar al desarrollo de una serie de conocimientos y procedimientos orientados a satisfacer dos grandes niveles de competencias: (1) resumir la información recogida, habitualmente cuantiosa, de un modo que resulte más comprensible y permita tomar decisiones útiles; y (2) inferir sobre una población numerosa en su tamaño, a partir de un subconjunto reducido de miembros de esa población.
- Ambas necesidades han dado lugar a las dos grandes ramas que tradicionalmente se suelen diferenciar dentro del campo de la Estadística, la estadística descriptiva y la estadística inferencial. La aplicación de ambas no es excluyente sino, con frecuencia, complementaria.
- Otra diferenciación también tradicional es la de estadística teórica versus estadística aplicada, la primera más dirigida al desarrollo y estudio de métodos formalmente válidos para hacer estadística, mientras que la segunda estaría orientada a la aplicación de esos métodos a campos de estudio concretos. El análisis de datos es un término que se usa habitualmente con el mismo sentido que el de estadística aplicada.

## 2. El método científico y la Estadística

- La Estadística se enmarca dentro del método que utiliza la Psicología en su desarrollo como Ciencia, el conocido como método científico.
- El método científico representa una estrategia ordenada y sistemática de actuación en la realización de un estudio o investigación. Es precisamente la aplicación del mismo, lo que permite dotar a una investigación del calificativo de científica. Las siguientes son características básicas inherentes al método científico:
  - (1) se aplica a cuestiones empíricamente contrastables;
  - (2) busca resultados de aplicación tan general como sea posible;
  - (3) da pie a la replicabilidad de los estudios basados en su aplicación;
  - (4) busca explicaciones parsimoniosas.
- Pasos que se pueden diferenciar en la aplicación del método científico (Navas, 2001):



**Ejemplo** ilustrativo de cómo se plasma la aplicación de algunas de las etapas del método científico en la búsqueda de una respuesta a la cuestión:

¿Cuál es la razón de que las personas difieran en su nivel de tolerancia al dolor?

- Formulación de una hipótesis:

Ejemplos de ésta en función del marco teórico de referencia considerado:

- \* Sociología/antropología: Las diferencias son función de factores culturales.
- \* Psicología social: Las diferencias son función de factores situacionales.
- \* Psicología de las diferencias individuales: Las diferencias son función de un rasgo de personalidad, la introversión-extraversión, de manera que los introvertidos tienen más tolerancia al dolor.

(En lo sucesivo, asumamos como nuestra la última de estas hipótesis)

- Definición operativa de las variables implicadas en la hipótesis: (→ asignatura de *Psicometría*)

- \* ¿De qué en concreto se va a recoger información?
- \* ¿Qué tipo de instrumento o estrategia vamos a utilizar a la hora de recoger esa información? (¿Cómo medir si alguien es extravertido o introvertido?; ¿cómo medir la tolerancia al dolor?)
- \* ¿Es fiable y válido el procedimiento que vamos a utilizar en la medida de las variables?

- Elección de un diseño de investigación:

- \* ¿De quién se va a recoger información? (diseño muestral)
- \* ¿Qué proceso lógico se va a seguir a la hora de recoger la información empírica que nos permita contrastar nuestra hipótesis? (diseño de recogida de los datos)

- Recogida de los datos

- Proceso y análisis de los datos:

Sea el caso de nuestro ejemplo en que, tras culminar la etapa del análisis de los datos, obtenemos el siguiente resumen de los datos recogidos en nuestra muestra (→ *Estadística descriptiva*):

Media tolerancia dolor introvertidos = 8

Media tolerancia dolor extravertidos = 6

(Supóngase que se ha medido la tolerancia al dolor en una escala de 1 [baja tolerancia] a 10 [alta tolerancia])

¿Podemos considerar que hay diferencias estadísticamente significativas entre estas medias a nivel poblacional (→ *Estadística inferencial*)?

- Interpretación de los resultados

- ¿Apoyan estos resultados la hipótesis inicial?
- ¿Qué consecuencias se derivan de los resultados obtenidos?

- En síntesis, la Estadística, como eslabón del proceder del método científico, nos va a permitir satisfacer el objetivo de resumir y transmitir de un modo comprensible la información procedente de datos empíricos (estadística descriptiva) así como, cuando sea oportuno, generalizar a partir de la información recogida de un conjunto reducido de sujetos a una población más amplia a la que éstos representen (estadística inferencial).

### 3. ¿Por qué la Estadística en el grado de Psicología?

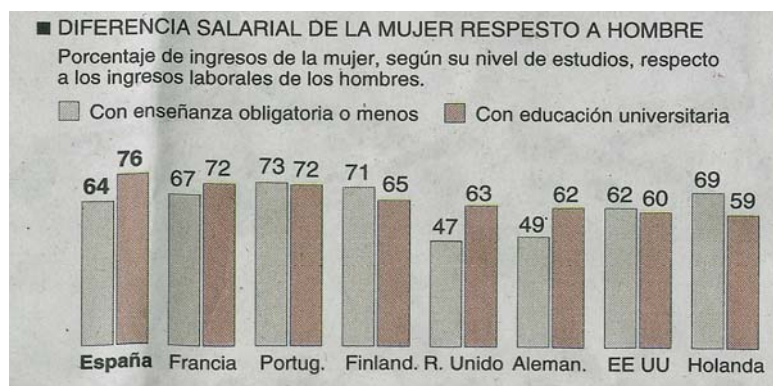
- Porque nos va a proporcionar un tipo de conocimientos y competencias que favorecen el pensamiento analítico y crítico.
- Porque nos va a capacitar para realizar estudios (investigaciones) en los que, bien poner a prueba conjeturas (hipótesis) que nos planteemos, bien buscar la respuesta a preguntas que nos surjan.
- Porque es crucial tener unos conocimientos básicos de Estadística para poder leer publicaciones (notas de prensa, artículos en revistas especializadas, informes de investigación, etc.) acerca de temas psicológicos, que son las que en el futuro van a permitirnos especializarnos y mejorar nuestro desempeño profesional.

A modo de **ejemplo** de lo último, aunque no se trate más que de un fragmento de los resultados de un artículo de investigación y de una representación gráfica resultante de otro estudio, debemos ser capaces de leer la información contenida en las mismas, interpretarla y, también, valorar la forma en que esa información ha sido elaborada y ofrecida a sus destinatarios.

1. “La comunicación familiar tiene un efecto directo y positivo en esta dimensión de la autoestima ( $\beta = .57$ ,  $p < .001$ ), que a su vez tiene un efecto negativo en el consumo de sustancias, mediado por el rechazo del adolescente a la autoridad ( $\beta = -.22$ ,  $p < .001$ ;  $\beta = .35$ ,  $p < .001$ ). Esta variable, el rechazo a la autoridad, es precisamente la que muestra un efecto directo más importante en el consumo de sustancias ( $\beta = .35$ ,  $p < .001$ ): un mayor rechazo a la autoridad institucional en los adolescentes influye en su mayor consumo de sustancias.”

*Fuente:* Cava, M. J., Murgui, S. y Musitu, G. (2008). Diferencias en factores de protección del consumo de sustancias en la adolescencia temprana y media. *Psicothema*, 20, 389-395.

2.



Fuente: Informe *Education at Glance* (OCDE, 2007)

#### 4. Algunos conceptos básicos de Estadística

- Un caso, entidad o unidad de observación es cada uno de los elementos sobre los que se desea recoger información en un determinado estudio. De forma sinónima son utilizados también con frecuencia los términos participante y sujeto. Ésta última denominación es apropiada cuando las unidades de observación son personas individuales, lo cual, aunque frecuente, no es siempre el caso: por una parte, las unidades de observación pueden ser díadas (p. ej., madre-hijo, parejas) o grupos (p. ej., familias, asociaciones, colegios, empresas); por otra parte, pueden ser animales (como es común en la investigación psicofisiológica) u objetos (p. ej., juguetes, anuncios radiofónicos...).
- Una variable es el conjunto de valores resultantes de medir una característica (propiedad, atributo) en las distintas unidades de observación objeto de estudio.
- Un dato (valor observado, observación) es un valor que proporciona información de un caso para una variable concreta.

**Ejercicio 1:** En un estudio con un grupo de 45 personas que habían sido tratadas psicológicamente por problemas de ansiedad, se preguntó a las mismas cuántas veces habían sufrido un ataque de ansiedad desde que terminó el tratamiento recibido. También se les preguntó si consideraban que había mejorado su calidad de vida a raíz del tratamiento psicológico que recibieron.

¿Cuántos casos y variables aparecen implicados en este estudio?, ¿cuántos datos se habrán obtenido en la recogida de datos?, ¿qué distintos valores adopta cada una de las variables (modalidades de cada variable)?

- Existen diferentes formas de organizar los datos recogidos en un estudio, los cuales es común que sean cuantiosos dado que es frecuente obtener datos de diversas variables para un conjunto amplio de casos. A estas formas de organización de los datos se les suele conocer como estructuras de datos, siendo la más utilizada en la práctica la conocida como tabla de datos (también denominada matriz de datos): se trata de una organización bidimensional en que las filas representan a las entidades y las columnas a las variables, constituyendo el cruce de una fila y una columna cualesquiera de la tabla, el valor observado (dato) correspondiente a un caso concreto en una determinada variable.

## Tipos de variables

Se han planteado diferentes clasificaciones de variables que varían en función del criterio considerado en su categorización. A continuación se presentan 3 de estas tipologías que resultan especialmente relevantes a la hora de comprender las implicaciones prácticas del concepto de variable.

(a) En función de cómo son medidas las variables, esto es, de la escala de medida utilizada en la obtención de los datos de la variable, se pueden diferenciar 3 tipos de variables:

- Variables cuantitativas o numéricas: aquéllas en que los valores resultantes de la medición son números que indican el grado o cantidad de lo que se está midiendo. Las variables cuantitativas se caracterizan por tener unidades de medida.

**Ejemplo:** la variable “Peso (gr.)” medida en una camada de 8 ratas ( $X$ ).

$X$ : {24,2; 39,6; 31,2; 27,8; 29,5; 36,5; 48,4; 42,0}

La unidad de medida en este caso es el gramo y los datos recogidos nos muestran, para cada rata, el número de unidades de medida que caracteriza a cada uno de ellas.

De las variables cuantitativas se hace una diferenciación adicional en función de que sean continuas o discretas. Las variables cuantitativas continuas pueden tomar cualquier valor numérico, esto es, entre cualquier par de valores, puede darse un valor numérico intermedio a aquellos dos. A modo de ejemplo, la variable “Tiempo empleado en completar una tarea orientada a evaluar la coordinación motriz”. Las variables cuantitativas discretas sólo pueden tomar ciertos valores concretos, por ejemplo, la variable “Número de hijos”.

En cualquier caso, hay que tener en cuenta que la precisión limitada de los instrumentos de medida provoca que, en la práctica, todas las variables cuantitativas sean medidas de un modo discreto aunque algunas, por su naturaleza, sean en realidad variables cuantitativas continuas.

- Variables cuasi-cuantitativas u ordinales: aquéllas en que los valores observados no son indicativos más que del orden o posición de las unidades de observación en lo que se esté midiendo. El dato correspondiente a un determinado caso tan sólo representa en que grado se es mayor o

menor que otro caso que tiene, respectivamente, un valor inferior o superior en aquello que se esté midiendo.

**Ejemplo:** la variable con los datos recogidos en un grupo de 121 personas a partir de la siguiente pregunta de un test: “Ansiedad que siente cuando se encuentra con mucha gente alrededor” ( $X$ ).

$X$ : {4, 2, 1, 2, 2, 3, 1, 3, 2, 4, 1, 3, 1, 2, 3, 2, 1, 1, 4, 1, 3...}

Las alternativas de respuesta a esta cuestión eran: Nada; Algo; Bastante; Mucha.

Codificación: [1: Nada; 2: Algo; 3: Bastante; 4: Mucha]

- Variables categóricas (cualitativas, nominales): aquéllas en que los valores no aportan ninguna información de magnitud ni de orden, tan sólo diferencian a los casos en distintas categorías de pertenencia. Una clasificación adicional de las variables categóricas diferencia a éstas entre dicotómicas (dos valores posibles) y politómicas (más de 2 valores posibles).

**Ejemplo:** la variable “Estado civil”, habiendo sido recogidos los datos para un total de 50 personas de la ciudad de Castellón ( $n = 50$ ):

$X$ : {0, 0, 1, 2, 2, 0, 1, 3, 2, 0, 1, 0, 1, 2, 0, 2, 1, 1, 0, 1, 0...}

Codificación: [0: soltero/a; 1: casado/a; 2: separado/a o divorciado/a; 3: viudo/a]

- A tener en cuenta en relación a los 3 tipos de variables diferenciados:
  - El tipo de variable (cuantitativa, cuasi-cuantitativa, categórica) es determinante en la selección del procedimiento estadístico a aplicar en su análisis.
  - A una característica (propiedad, atributo) que en su medición adopta los mismos valores para todas las entidades se le conoce como una constante. No se habla en este caso de variable dado que falta la variabilidad en los valores observados inherente a este término. Destacar que la medición de una misma característica puede dar lugar a una variable o a una constante en función del estudio de que se trate (p. ej., el ‘Sexo’ en un estudio sobre la depresión post-parto y en un estudio sobre la ansiedad).
  - Una variable no es de un tipo u otro *per se*, sino que va a depender del modo en que se lleve a cabo su medición (p. ej., la edad o el consumo de tabaco pueden ser medidas utilizando diferentes escalas de medida que den lugar a variables de diferente tipo).



**Ejercicio 2:** ¿De qué tipo son las siguientes variables en función de la escala de medida?

- Tiempo en segundos invertido en recorrer un laberinto.
- Memoria visual medida a través del número de caras reconocidas en una serie de 100 imágenes de caras, de las que 50 habían sido presentadas una hora antes.
- Nivel de ingesta alcohólica medida como muy alta, alta, media, baja y nula.
- Clasificación de un grupo de sujetos en función de su lugar de residencia (urbana, rural).
- Amigabilidad de los niños medida como la ordenación de los mismos realizada por su profesora.
- Inteligencia de un grupo de niños medida a partir de la aplicación del WISC.
- Dosis de un fármaco medida en mililitros.
- Grado de invalidez (33, 66, total)
- Tipo de colegio al que se asiste (público, privado, concertado)
- Clasificación hecha por una psicóloga de un grupo de pacientes que realizan terapia de grupo en función del grado de mejoría de los mismos.
- Nacionalidad.
- Seguridad laboral de las empresas medida como el número de accidentes laborales ocurridos en la empresa durante el último año.
- Grupo al que se pertenece en la realización de una investigación (tratamiento/control).
- Número de miembros en la unidad familiar.
- Rendimiento académico medido como el número de respuestas correctas en un examen.
- Clasificación de películas en función de su género (comedia, terror...)
- Estado de ánimo (positivo, neutro, negativo)

**(b)** En función de qué se mide:

Esta tipología diferencia entre aquellas variables que son propiamente el objeto de interés de nuestro campo de conocimiento de aquéllas que, aunque no de naturaleza psicológica, suelen ser consideradas en un gran número de investigaciones psicológicas dado el interés que suele tener plantear análisis diferenciales en función de las mismas.

- Variables psicológicas: son variables que hacen referencia a características relativas a la personalidad, la inteligencia, conocimientos, hábitos, aptitudes, actitudes y habilidades, entre otras facetas de la psicología.

- Variables socio-demográficas: aquéllas que hacen referencia a aspectos biológicos y sociológicos tales como la edad, el sexo, el lugar de residencia, el nivel de estudios alcanzado, el estado civil...

(c) En función de qué rol desempeñan en el diseño de la investigación:

- En bastantes estudios se plantean hipótesis sobre la relación entre 2 variables (hipótesis bivariadas) y, más concretamente, sobre la existencia de influencia (o efecto) de una variable sobre otra -por ejemplo, en una investigación se podría hipotetizar la influencia de la inteligencia emocional sobre la conducta solidaria. En estos casos se habla de una variable explicativa que se conjetura que es la causa de una variable de respuesta. En la literatura aparece cierta diversidad en la forma en que estos dos tipos de variables son nombrados, tal como se pone de manifiesto en la siguiente tabla:

Variable explicativa		Variable de respuesta
Variable independiente (VI)	⇒	Variable dependiente (VD)
Variable predictora		Variable criterio
Variable X		Variable Y

- Cabe destacar que una misma variable puede aparecer como explicativa en un estudio, mientras que variable de respuesta en otro diferente, por ejemplo, una variable como la inteligencia emocional es la variable explicativa en el ejemplo anterior de la conducta solidaria, mientras que sería la variable de respuesta en un estudio en que se hipotetizara el efecto del estilo educativo que se ha recibido sobre la inteligencia emocional.

- Si en una hipótesis aparecen implicadas más de 2 variables, se hace referencia a la misma como hipótesis multivariada. En este caso, aunque se pueden dar diferentes tipos de combinaciones, es el caso más común encontrarse que una de las variables adopte el papel de variable de respuesta, mientras que el resto sean variables explicativas.

- Es frecuente que en estudios en que se desea analizar la influencia de una variable sobre otra, los valores de la primera, la variable explicativa, sean asignados a los sujetos por parte de quien lleva a cabo el estudio. A este caso de variable explicativa en que, en realidad, no hay un proceso de medición a partir del que se obtengan los datos, sino que es el responsable de la investigación quien determina los valores que tendrán los sujetos en esa variable, se le conoce como variable manipulada.

**Ejemplo:** en un estudio en que se quiera estudiar el efecto de diferentes dosis de una determinada sustancia psicoactiva sobre la conducta, una estrategia habitual es que se apliquen diferentes dosis de la misma, establecidas a priori por el investigador (por ejemplo, 100, 200 y 300 mg.), a diferentes grupos de sujetos. Así, la variable “Dosis de la sustancia administrada” sería una variable manipulada. Otro ejemplo: la variable “Grupo en el que se participa en un estudio”, en que se quiere comparar la eficacia de una determinada terapia psicológica (Grupo A) frente a otra (Grupo B). Los sujetos son asignados a un grupo u otro por el investigador, no es algo que los sujetos decidan o sean ya a priori.

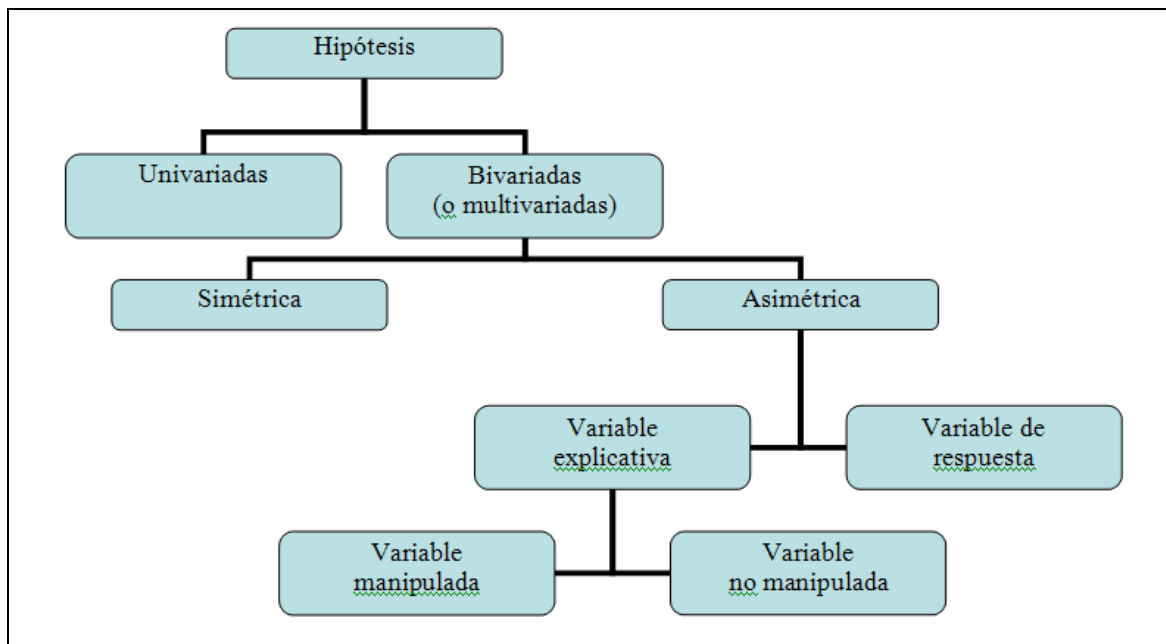
- También podemos encontrar estudios en que se planteen hipótesis sobre la relación entre variables sin que esté claro a priori que una variable sea explicativa y la otra variable de respuesta. En estos casos se habla de relación simétrica entre las variables, frente a la relación asimétrica que se daba en los ejemplos previos.

**Ejemplo** de relación simétrica: un estudio en que aparecieran implicadas las variables “Autoestima” y “Habilidades sociales”, caso en el que no hay relación de causalidad entre ambas que esté clara a priori (puede darse en los dos sentidos).

- Del mismo modo que no es aplicable la distinción entre variable explicativa y variable de respuesta si nos encontramos con una relación simétrica entre variables, tampoco lo será en aquellos estudios que implican a una única variable (*hipótesis univariadas*).

**Ejemplo:** Un estudio en que se quiera comprobar la hipótesis de que la mayoría de jóvenes españoles de entre 25 y 30 están solteros, implica la obtención de datos de una única variable, el “Estado civil”, por lo que en este estudio concreto no tiene sentido hablar de que sea variable predictora o criterio de ninguna otra variable.

- A modo de síntesis gráfica:



**Ejercicio 3:** Para cada una de las siguientes hipótesis:

- ❑ Aplica el árbol de decisión superior a fin de definir de qué tipo son las siguientes hipótesis y las variables implicadas en las mismas (excepto la rama de variable manipulada/no man.).
  - ❑ Cuando haya información suficiente (o se pueda suponer), señalar el tipo de variable en función de la escala de medida.
- Entre los estudiantes que abandonan la Universidad, la proporción de hombres y mujeres es muy similar.
  - En la población española de niños/as entre 3 y 10 años, la hiperactividad es más frecuente en niños que en niñas.
  - La prevalencia de los trastornos mentales es superior en zonas urbanas que en zonas rurales.
  - Más del 50% de los conflictos agresivos entre niños en edad preescolar se relacionan con la posesión de un objeto.
  - En delincuentes juveniles, la gravedad de los delitos cometidos es inversamente proporcional a la edad en que comenzaron a delinquir.
  - A los 4 años, la mayoría de los niños son capaces de nombrar los colores primarios.
  - El nivel educativo se relaciona con la aparición de demencia en la 3ª edad.
  - Para los jóvenes europeos, el paro es el problema social más importante.
  - Los hijos de padres casados son más estables emocionalmente que los hijos de padres separados y, estos últimos, más que los hijos de padres solteros.

- La enseñanza multimedia del Inglés en ESO es más efectiva que la enseñanza tradicional.
- La sobrecarga laboral está relacionada con la aparición del síndrome de Burnout.
- El absentismo laboral es una consecuencia relevante del síndrome de Burnout.
- 2 de cada 3 fumadores en España quieren dejar de fumar.
- Entorno al 10% de los trabajadores europeos del sector de servicios ha sufrido de acoso psicológico en el trabajo (mobbing) durante el último año.
- Existe una relación positiva entre bienestar psicológico y longevidad.
- La proporción de votantes del PSOE en las próximas elecciones generales en España será  $\geq 0.50$ .
- La integración de los niños inmigrantes en los colegios de educación primaria es más óptima en medios rurales que en medios urbanos.
- El bienestar psicológico de los ancianos que viven con sus familias es superior al de los que viven en residencias geriátricas.
- En niños autistas la prevalencia de conductas agresivas es del 20%.
- Las calificaciones obtenidas en la asignatura de Estadística están directamente relacionadas con las obtenidas en Fundamentos de Psicobiología.
- El nuevo fármaco (A) contra la ansiedad ofrece mejores resultados que los fármacos tradicionales (B) y (C).

## 5. Metodologías de investigación y estadística

- Se pueden plantear diferentes clasificaciones de las metodologías o diseños de investigación utilizados en la práctica. Una clásica es la que diferencia 3 grandes categorías que varían en el nivel de control interno aplicado en el diseño de la investigación. Cuanto mayor sea ese control interno (grado de intervención en la forma de llevar a cabo el estudio), con mayor seguridad se puede llegar a afirmar que las diferencias de los sujetos en la variable de respuesta se deben a la variable(s) explicativa(s) y no a otras variables –en terminología científica se habla de una mayor validez interna. Sin embargo, ese mayor control interno suele venir asociado a mayores dificultades a la hora de llevarlo a cabo un estudio en la práctica, así como a situaciones más artificiales y alejadas de la realidad y, por lo tanto, con una menor representatividad del contexto sobre el que normalmente queremos generalizar las conclusiones –en terminología científica se utiliza el término validez externa para hacer referencia a esta faceta del grado de realismo o representatividad de un

estudio. Es deseable que cualquier investigación tenga tanto alta validez interna como externa, si bien, es frecuente que ambos aspectos se contrapongan en su consecución en la práctica.

• Clasificación de las metodologías de investigación:

- Metodología experimental:

- La variable explicativa o independiente es manipulada por el investigador/a, quién va a definir a priori los valores que ésta puede tomar (→ variable manipulada).
- Asignación aleatoria de los participantes en el estudio a los diferentes subgrupos definidos por los valores de la variable explicativa.

- Metodología cuasi-experimental:

- Manipulación de la variable explicativa (idem a la metodología experimental).
- Asignación no aleatoria, sino determinada por condicionantes prácticos como, por ejemplo, el aprovechamiento de la existencia de grupos naturales a los que asignar los diferentes valores de la variable explicativa.

- Metodología no experimental:

- No hay manipulación de la variable explicativa, los valores de ésta vienen ya dados.
- Dentro de esta categoría se suelen encuadrar dos tipos de diseños de investigación ampliamente utilizados en el ámbito de la investigación en las Ciencias del Comportamiento: la metodología observacional y la metodología selectiva (también denominada correlacional).

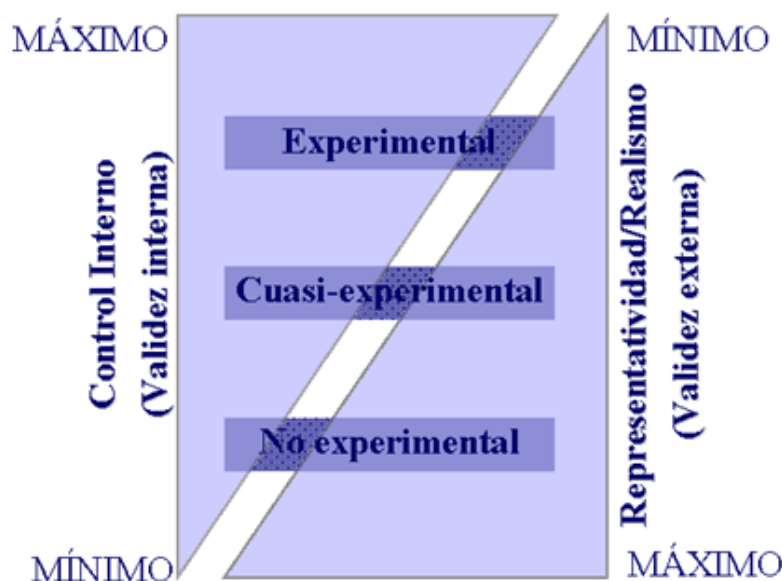


Figura resumen de las metodologías de investigación (adaptado de Portell, Vives y Boixadós, 2003)

- Subrayar que podemos encontrar bastantes ejemplos de estudios en el ámbito de la investigación psicológica en los que no es posible manipular la variable explicativa, por ejemplo, en todos aquellos en que ésta sea alguna característica innata o ya adquirida por los sujetos. Supóngase el estudio del efecto del optimismo/pesimismo sobre la capacidad de recuperación de una enfermedad: no se le puede decir a un grupo de sujetos que sea optimista y al otro pesimista...

- Aunque no es siempre posible, en algunos estudios se puede optar por la aplicación de cualquiera de estas tres metodologías. Será el responsable del diseño de la investigación quien, en función de los recursos disponibles y del tipo de conclusiones a que se quiera llegar, decida qué tipo de diseño de investigación implementar. Supóngase que se quiere comparar la eficiencia de la aproximación farmacológica frente a la psicológica en el tratamiento de la ansiedad:

- **Ejemplo** de diseño de investigación experimental → Se dispone de un grupo de sujetos con problemas de ansiedad que desean recibir algún tipo de tratamiento. Se decide crear dos subgrupos, uno que recibirá tratamiento farmacológico y otro que lo recibirá psicológico. Los sujetos son asignados aleatoriamente, la mitad de ellos a un grupo y la otra mitad al otro. Se controla de forma cuidadosa que a lo largo de la aplicación del tratamiento no haya otras variables que puedan influir de forma diferencial a un grupo y a otro y que, por lo tanto, puedan provocar divergencias en ambos grupos. Tras finalizar el tratamiento, se aplica un test a los sujetos de ambos grupos para evaluar el estado de su trastorno de ansiedad, siendo las puntuaciones en éste las que servirán de base para compararlos y extraer conclusiones.

- **Ejemplo** de diseño de investigación cuasi-experimental → Se decide crear dos subgrupos, uno que recibirá tratamiento farmacológico y un segundo que lo recibirá psicológico. Al primero son asignados un conjunto de sujetos que han acudido a la sección de psiquiatría de un hospital a recibir tratamiento, mientras que al segundo los procedentes de una clínica psicológica. Ídem al anterior en lo que respecta al control de variables extrañas y en lo relativo a la medida de la variable de respuesta.

- **Ejemplo** de diseño de investigación no experimental → En una serie de clínicas se recoge información de sujetos que hayan recibido tratamiento para la ansiedad, ya sea psicológico o farmacológico, y que hubiesen pasado después una determinado test de ansiedad para evaluar su estado. Se comparan las puntuaciones en ese test de aquéllos que han recibido tratamiento psicológico frente a los que lo han recibido de carácter farmacológico.

• Diseño de investigación y estadística: el análisis estadístico de los datos derivados de un estudio va a permitir obtener los resultados a partir de los que extraer las conclusiones oportunas. Ahora bien, debe resaltarse que la validez de esas conclusiones va a depender no sólo del análisis de los datos sino también, y muy importante, del diseño de investigación por el que se hubiese optado a la hora de llevar a cabo el estudio.

Siguiendo con nuestro **ejemplo**, supóngase que se han obtenido en nuestro estudio sobre el tratamiento de la ansiedad los datos que se muestran a continuación (tipo de diseño de investigación: cualquiera de ellos):

<u>Trat_F</u>	<u>Trat_P</u>
6	4
8	9
11	4
...	...
<u><math>\bar{X}=8</math></u>	<u><math>\bar{X}=6</math></u>

Tras realizar el oportuno análisis de los datos recogidos (obtención de la media de ambos grupos), puede observarse como la media del grupo del tratamiento farmacológico es mayor que la del tratamiento psicológico y, asumiendo que mayor puntuación en el test representa un peor resultado (mayor nivel de ansiedad), los resultados sugieren concluir que el tratamiento psicológico es más eficaz que el farmacológico. Ahora bien, ¿cuál es la validez de esa conclusión?, ¿se puede concluir que la causa de la diferencia en ansiedad entre ambos grupos se debe al tratamiento? La respuesta a estas cuestiones no viene determinada por el análisis de los datos, sino por el diseño de investigación que se hubiese aplicado.

En el caso que se hubiese aplicado un diseño no experimental, tan sólo podremos concluir que existe relación entre la variable “Tratamiento” y la variable “Ansiedad”, en concreto, que al tratamiento psicológico aparecen asociados niveles más bajos de ansiedad, mientras que lo contrario sucede con el tratamiento farmacológico. Ir más allá, definir una relación causal de la variable “Tratamiento” sobre la variable “Ansiedad”, poder concluir que el tipo de tratamiento es la causa de las diferencias encontradas en ansiedad en ambos grupos, sólo será posible si se ha aplicado un diseño de investigación experimental. Por último, en el caso de aplicar un diseño cuasi-experimental, la conclusión acerca de la relación causal entre ambas variables estará matizada por la incertidumbre, menor cuanto mejor se justifique la equivalencia de los grupos en el momento de iniciarse el tratamiento.



**Ejercicio 4:** Determinar cuál sería la metodología más adecuada para contrastar empíricamente cada una de las hipótesis bivariadas del ejercicio 3 (una pista: pensar si la variable explicativa se puede manipular o no).

## 6. Estadística descriptiva y estadística inferencial

- Una diferenciación tradicional en el campo de la estadística ha sido la que distingue entre, por una parte, el interés de esta disciplina por resumir los datos recogidos de una forma que resulte informativa, comprensible y permita tomar decisiones útiles (estadística descriptiva) y, por otra parte, el interés por inferir sobre una población numerosa en su tamaño, a partir de un subconjunto reducido de miembros de esa población (estadística inferencial). En la práctica, la aplicación de ambas no es excluyente sino, con frecuencia, complementaria.

Asociada a esta distinción en el campo de la estadística se encuentra la diferenciación entre los pares de conceptos que a continuación se presentan.

### 6.1. Población y muestra

- “Se llama población estadística al conjunto de todos los elementos que cumplen una o varias características o propiedades” (Botella et al., 2001).

Otra propuesta de definición: Conjunto de entidades objeto de interés en un estudio.

- En cuanto que conjunto, sus elementos tendrán alguna característica(s) en común que es la que va a determinar su pertenencia a ese conjunto. La definición de la población en un estudio debe expresar con precisión esas características, pues éstas representan el criterio de pertenencia a la misma, permitiendo discernir con claridad quién y quién no forma parte de la población objeto de estudio.

**Ejemplos** de poblaciones:

- La población de mujeres de entre 25 y 35 años de la CV
  - Criterios de pertenencia: ser mujer, tener entre 25 y 35 años y ser de la CV.
  - Posibles ambigüedades: entre 25 y 35 años, ¿ambos inclusive?; ser de la CV, ¿vivir?
- Los estudiantes del grupo de 1º D de la Facultad de Psicología de la UVEG del presente curso académico.

- Los colegios privados o concertados de la ciudad de Valencia.

- La población que se especifique como objeto de estudio va a determinar de quién se va a recoger datos (ya sea de todos sus elementos o de una parte de ellos) y, también, sobre quién van a recaer las conclusiones derivadas del estudio. Es por ello que su especificación representa un aspecto crucial en el diseño de cualquier investigación.
- En la difusión de los resultados de un estudio deben detallarse las características de la población objeto de estudio. Ello no es siempre lo habitual en las noticias sobre estudios o investigaciones difundidas en medios de comunicación de masas (prensa, radio, TV...). En algunos casos, ese ocultamiento u omisión está motivado por la premura de espacio y tiempo que estos medios tienen para informar, pero esta premura acaba convirtiéndose a veces en una excusa para omitir o esconder información de forma malintencionada.

Los oyentes o lectores expertos serán escépticos ante un informe en que se omita esa información, pero no ocurrirá lo mismo con otros receptores no expertos, que pueden creer que los resultados hacen referencia a una población amplia cuando, en realidad, puede que no sea éste el caso. Es un deber ético el ser transparente en un aspecto tan crucial como es el de la descripción de las características de la población objeto de estudio y las de los participantes de los que se haya recogido datos, pues ello va a permitir valorar la capacidad de generalización de los resultados obtenidos con éstos a aquélla.

- Algunas posibles fuentes de confusión acerca del concepto de población:
  - Audiencias no expertas pueden entender el concepto estadístico de población en su acepción geográfica, en cuanto que pueblo, ciudad o similar.
  - Que una población sea concebida como un conjunto necesariamente numeroso de elementos. A modo de **ejemplo** de ello: “Precisamente el hecho de que las poblaciones, por lo general, sean infinitas o estén formadas por un gran número de elementos, hace que la descripción exacta de sus propiedades sea un objetivo prácticamente inaccesible.” (Pardo y San Martín, 2001).

El tamaño de la población dependerá de cuál es el objetivo del estudio que se plantee llevar a cabo y, más específicamente, de sobre quién se desee extraer conclusiones. Por **ejemplo**, el tamaño de la población puede ser bien reducido en el estudio de la eficacia de un método determinado de lecto-escritura que tenga como población de interés simplemente a los

escolares de una aula de un colegio concreto, como podría ser el caso de un estudio en un colegio en el que estemos desarrollando nuestro trabajo como psicólogos.

- “Una muestra es un subconjunto de los elementos de una población” (Botella et al., 2001).

Tal vez se podría añadir: ...sobre el que obtener la información que nos permita inferir alguna característica(s) de esa población.

- Ventajas de trabajar con muestras en vez de con las respectivas poblaciones:

- Economía de recursos: obviamente, cuanto menor el número de casos de los que haya que recoger datos, menor será el trabajo implicado en ello.
- Calidad de los datos recogidos.

- Una ventaja mal entendida que se atribuye a trabajar con una muestra, frente a hacerlo con la población correspondiente, es la derivada del problema de accesibilidad a los elementos de la población.

Por **ejemplo**: “Las poblaciones que habitualmente interesa estudiar en Psicología o son infinitas o son tan grandes que normalmente resulta muy difícil (si no imposible) tener acceso a todos sus elementos. Bajo estas circunstancias, es de las muestras de donde podemos obtener la información necesaria...” (Pardo y San Martín, 2001).

En realidad, el problema no sería tanto la gran dificultad, o imposibilidad, para tener acceso a los elementos de una población numerosa -a no ser que se trate de poblaciones que no tengan una localización tangible en el espacio o en el tiempo-, sino más bien lo costoso que puede resultar el acceder a los mismos.

- Análogamente a lo que ocurre con el concepto de población, una fuente de confusión bastante frecuente al considerar el concepto de muestra es asociarlo a un conjunto de elementos de tamaño reducido.

**Ejemplo**: “Al contrario de lo que ocurre con las poblaciones, que suelen ser conjuntos de elementos de gran tamaño, las muestras suelen ser conjuntos de elementos de tamaño reducido.” (Pardo y San Martín, 2001).

Una buena muestra de una población numerosa puede tener un tamaño considerable. Así, la población del ejemplo anterior de una aula de un colegio podría limitarse a tan solo 30 casos, mientras que la muestra correspondiente a un estudio en que la población sean los habitantes de un país puede perfectamente tener un tamaño de tres o cuatro mil casos.

## 6.2. Parámetros y estadísticos

- “Un parámetro es un valor numérico que describe una característica de una población” (Botella et al., 2001).

Otra propuesta de definición: Es el valor de un índice estadístico (media, varianza, proporción...) obtenido a partir de los datos de una población, que describe alguna característica de la misma.

- “Un estadístico es un valor numérico que describe una característica de una muestra” (Botella et al., 2001)

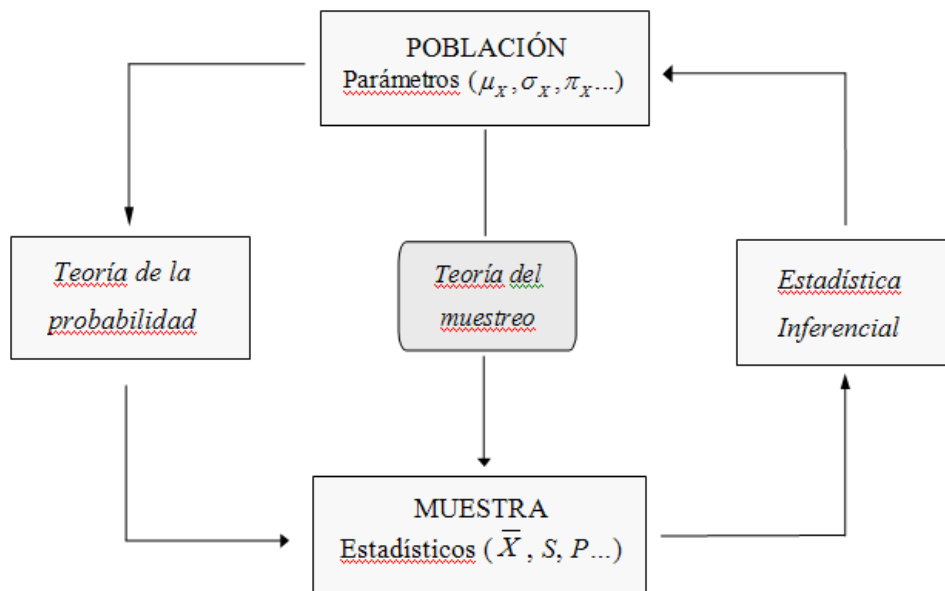
Otra propuesta de definición: Es el valor de un índice (estadístico) obtenido a partir de los datos de una muestra que nos va a permitir inferir el valor de alguna característica de la población a la que esa muestra representa.

**Ejemplo:** el índice estadístico de la media aritmética  $\left(\frac{\sum x_i}{n}\right)$ , si lo obtenemos a partir de los datos de una muestra obtendremos un estadístico, mientras que si a partir de los de una población, un parámetro.

- Mientras que el valor obtenido para un parámetro es una constante, el obtenido para un estadístico se puede concebir como una variable, dado que para diferentes muestras de una misma población se obtendrán normalmente diferentes valores.
- Es costumbre representar simbólicamente a los parámetros con letras griegas minúsculas ( $\mu_x, \sigma_x, \pi_x \dots$ ), mientras que a los estadísticos con letras latinas mayúsculas ( $\bar{X}, S_x, P_x \dots$ ).

• Relacionando conceptos:

	<i>Muestra</i>	<i>Población</i>
<i>Est. Descriptiva</i>	(→ Estadísticos)	(→ Parámetros)
<i>Est. Inferencial</i>	(→ Estimaciones de parámetros)	×



**Ejercicio 5:** Para cada uno de los siguientes estudios:

- Identificar la población objeto de estudio en cada caso y diferenciar si se ha optado por trabajar con una muestra de esa población o con la población.
  - Identificar si el estudio sería únicamente descriptivo, o bien, descriptivo e inferencial.
- Un empresario quiere conocer el nivel de satisfacción laboral de los 20 empleados de su empresa. Para ello les pasa un cuestionario preguntándoles por su satisfacción respecto de diferentes aspectos de su trabajo.
  - A partir de una muestra de mujeres embarazadas, nos disponemos a estudiar la relación entre el apoyo social recibido por la mujer durante el embarazo y el estado de salud del bebé al nacer (medido a partir de la puntuación en el test de APGAR).
  - Un psicólogo de un centro educativo desea aumentar la inteligencia emocional de los niños de primaria de dicho centro. Para ello aplica un programa de intervención y para evaluar si ha sido efectivo, compara el grado de conflictividad de las interacciones de estos niños con otros antes y después de la aplicación del programa.

- Queremos conocer la incidencia del estrés laboral en la población de trabajadores de la rama de hostelería de la CV. Para ello seleccionamos al azar un grupo de empresas de la CV de este sector y encuestamos a sus trabajadores.
- Un psicólogo de servicios sociales de un ayuntamiento quiere conocer la actitud hacia la inmigración de los habitantes de un barrio conflictivo de la ciudad. Para ello realiza una encuesta telefónica a 100 personas censadas en dicho barrio.
- Un psicólogo de una empresa nacional quiere evaluar si la implantación de un nuevo sistema de calidad en la producción ha aumentado la productividad de los trabajadores de dicha empresa. Para ello compara el número de unidades fabricadas antes y después de la implantación del sistema de calidad en un subconjunto de centros de dicha empresa.
- Un psicólogo escolar quiere conocer la satisfacción de los padres de los niños del colegio con las actividades extraescolares realizadas por sus hijos. Para ello convoca a los padres a una reunión y les pasa un cuestionario al respecto.
- Un psicólogo clínico quiere evaluar la eficacia de una terapia de tipo conductual aplicada a un paciente con claustrofobia. Para ello, tras la aplicación de la terapia, mide su nivel de ansiedad al estar dentro de un ascensor y lo compara con el obtenido antes de iniciarse la terapia.
- Para estudiar si la integración de los niños inmigrantes en los colegios de educación primaria es más óptima en medios rurales que en medios urbanos se seleccionan al azar dos colegios (uno rural y otro urbano) y se evalúa mediante cuestionarios a los profesores el nivel de integración de los niños inmigrantes en las aulas.

## Referencias:

- Aaron, A. y Aaron, E. N. (2001). Estadística para Psicología. Buenos Aires: Prentice Hall.
- Botella, J., León, O. G., San Martín, R. y Barriopedro, M. I. (2001). Análisis de datos en psicología I: teoría y ejercicios. Madrid: Pirámide.
- Gambara, H. (2002). Métodos de investigación en psicología y educación: Cuaderno de prácticas (3ª ed.). Madrid: McGraw-Hill.
- Navas, M.J. (2001). Métodos, diseños y técnicas de investigación psicológica. Madrid: UNED.
- Pardo, A. y San Martín, R. (2001). Análisis de datos en psicología II. Madrid: Pirámide.
- Portell, M.; Vives, J. y Boixadós, M. (2003). Mètodes d'investigació: recursos didàctics. Bellaterra: Servei de Publicacions de la UAB.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.