

# INVESTIGAR EN EDUCACIÓN

(Universidad de Encarnación, Paraguay, junio 2018)

Francesc J. Hernández (Universitat de València)

## INTRODUCCIÓN

### I.- SOBRE LAS TÉCNICAS DE INVESTIGACIÓN CUANTITATIVAS

- 1.- DEFINICIONES PRELIMINARES
- 2.- UN EJEMPLO INTRODUCTORIO
- 3.- ENCUESTAS
  - 3.1.- Sobre la elaboración de cuestionarios
  - 3.2.- Encuestas y comparaciones
  - 3.3.- Atención a «ns/nc»
  - 3.4.- Un ejemplo de autoevaluación
- 4.- ÍNDICES Y FÓRMULAS
  - 4.1.- Índices
  - 4.2.- Fórmulas
- 5.- RELACIÓN DE DOS VARIABLES
  - 5.1.- Determinación de tendencias
  - 5.2.- Líneas de tendencia
  - 5.3.- Coeficientes de correlación
  - 5.4.- Interpretación de R y de R<sup>2</sup>
  - 5.5.- Depuración de datos y elección de la línea de tendencia
- 6.- RELACIÓN DE MÁS DE DOS VARIABLES
  - 6.1.- Comparación de correlaciones
  - 6.2.- Matrices de correlaciones
  - 6.3.- Análisis factorial

### II.- SOBRE LAS TÉCNICAS DE INVESTIGACIÓN CUALITATIVAS

- 1.- DEFINICIONES PRELIMINARES
- 2.- ENTREVISTAS A INFORMADORES PRIVILEGIADOS
  - 2.1.- Antes de la entrevista
  - 2.2.- Durante la entrevista
  - 2.3.- Después de la entrevista
- 3.- ENTREVISTAS A PERSONAS EN GENERAL
4. GRUPOS DE DISCUSIÓN
- 5.- TRANSCRIPCIÓN E INTERPRETACIÓN
  - 5.1.- Transcripción
  - 5.2.- Interpretación
- 6.- INVESTIGACIÓN BIOGRÁFICA

## INTRODUCCIÓN

---

En ciencias humanas y sociales se usan técnicas de investigación cuantitativas y cualitativas. En general podríamos decir que ambos tipos de técnicas se basan en la *comparación*, aunque las técnicas cuantitativas pretenden realizar una *medición* y las técnicas cualitativas una *comprensión*.

Hay muchos manuales que explican el uso de técnicas de investigación cuantitativas y cualitativas referidas al ámbito educativo. El objetivo de este texto no es sustituir a los manuales, sino ofrecer consejos u observaciones prácticas del uso de técnicas, ilustrados generalmente con ejemplos extraídos de proyectos de investigación del autor.

A continuación se realizarán algunas clarificaciones conceptuales y se comentarán errores frecuentes.

La investigación científica tiene como objetivo establecer teorías científicas o someterlas a evaluación. Las teorías científicas se expresan en enunciados no particulares, es decir, aplicables a una generalidad de casos, que establecen nexos de necesidad (o probabilidad) entre estados de cosas. Es decir, la investigación científica no pretende saber más de *un caso*, sino formular enunciados generales, del tipo: «siempre que sucede A, pasará B, por la razón C» o «si sucede A, habrá una probabilidad P de que pase B, por la razón C».

La forma matemática de una ley científica suele ser una ecuación. Tómese como ejemplo el tercer principio de la dinámica de Newton:

$$f = m \cdot a$$

(fuerza es igual a masa por aceleración)

Significa que entre la variable fuerza y la variable aceleración hay una relación de proporcionalidad directa:  $f \propto a$ , que podemos transformar en una ecuación utilizando una constante (en este caso, m).

En las ciencias humanas y sociales es muy difícil establecer leyes (universales y necesarias). Por ello, muchas veces hemos de conformarnos con teoremas o con conjeturas.

Un teorema es una verdad demostrable, pero que no tiene el rango de una ley. Por ejemplo, en el contexto europeo se constata provisionalmente que:

$$D^2 \propto E \cdot A$$

Donde D es una medida de desigualdad social (como, por ejemplo, la relación de ingresos entre el quintil superior e inferior; entre el decil superior e inferior o el

índice de Gini de desigualdad), E es la expectativa de escolarización (expresada en años) y A es la proporción de abandono educativo (no confundir con abandono escolar, más adelante se precisan ambos términos). De manera que podríamos establecer que:

$$D^2 = E.A.k$$

siendo k una constante para armonizar las diferentes unidades de medida.

(Más adelante se explicará cómo justificar una relación de proporcionalidad a partir de coeficientes de correlación).

Ahora bien, se precisan más estudios para demostrar este teorema y encajarlo en un cuerpo teórico que explique la relación entre desigualdad y abandono.

De un teorema podemos extraer corolarios.

Una conjetura es una suposición verosímil que permite orientar la investigación científica. Por ejemplo, después de realizar muchas entrevistas podemos llegar a la afirmación siguiente: los estudiantes de clase inferior que acceden a la universidad se matriculan en aquellas carreras cuyos estudios les complacen; los de la clase superior se matriculan en las carreras función de la profesión que desean desempeñar.

El hecho de que sea muy difícil ofrecer una demostración completa de una conjetura como la citada anteriormente no significa que no sea útil para el desarrollo de la investigación.

Una hipótesis no es una conjetura (ya que esta tiene verosimilitud y probablemente no pueda ser demostrada satisfactoriamente), sino una suposición que orienta un proceso de investigación. A diferencia de una mera suposición, una hipótesis debe estar vinculada con un contexto de demostración, es decir, tenemos que saber qué hechos verificaran o refutaran la hipótesis.

En conclusión, en la investigación científica en general y aquella referida a la educación en particular se deben perseguir enunciados generales y siempre hacer constar si nos encontramos ante leyes, teoremas, corolarios, conjeturas o hipótesis.

El prestigioso periódico británico *The Times* publicó en agosto de 2011 un reportaje sobre el consumo de Coca-cola, con un gráfico donde se representaban el consumo anual per cápita de esa bebida en diversos países europeos. Tal vez por la afición británica a su sistema de medidas imperiales, la cantidad de líquido consumida no se expresaba en litros, sino en el número equivalente de los clásicos botellines de 6,5 onzas. Según el gráfico, en Estados Unidos el promedio era de 394 botellines por persona y año. En Malta el promedio era de 606 botellines. ¿Por qué los habitantes de la isla mediterránea consumían un 50% más de refresco que los habitantes del país donde se inventó la bebida? ¿A qué se debe el incremento que hace de ellos líderes mundiales del consumo del líquido refrescante?

Es fácil suponer que no se trata de ninguna pasión maltesa por el brebaje de Pemberton, sino simplemente de un cálculo erróneo, probablemente derivado de una confusión en las personas que habían cubierto la información, ya sea periodistas o responsables del grafismo, en el dato de la población de hecho y la población de derecho de la isla de Malta. Si en lugar de considerar las personas que efectivamente se encuentran en la isla (población de hecho), muchas de ellas turistas, se considera la cantidad de la ciudadanía maltesa (población de derecho), el porcentaje por persona de botellines de Coca-cola naturalmente se incrementa.

Los errores en la investigación en general y la investigación educativa en particular pueden provenir de diversas fuentes.

En algunos casos se trata de confusiones, como en el ejemplo citado cuando se usa la población de derecho en lugar de la población de hecho. En los estudios sobre la educación se cometen a veces errores así:

a) Cuando se confunde «abandono escolar» (alumnado que no concluye el ciclo obligatorio) y «abandono educativo» (alumnado que no cursa una secundaria postobligatoria).

b) Cuando se confunde «educación superior» (terciaria) y «educación universitaria».

Para no incurrir en estos errores se puede apelar la clasificación ISCED de UNESCO

Nivel 0.- Educación Infantil	Level 0.- Early childhood education
Nivel 1.- Educación Primaria	Level 1.- Primary
Nivel 2.- Educación Secundaria obligatoria	Level 2.- Lower secondary
Nivel 3.- Educación Secundaria postobligatoria	Level 3.- Upper secondary
Nivel 4.- Educación postsecundaria no terciaria	Level 4.- Post-secondary non-tertiary
Nivel 5.- Educación terciaria de ciclo corto	Level 5.- Short-cycle tertiary
Nivel 6.- Grado o equivalente	Level 6.- Bachelor or equivalent
Nivel 7.- Máster o equivalente	Level 7.- Master or equivalent
Nivel 8.- Doctorado o equivalente	Level 8.- Doctoral or equivalent

Fuente: United Nations Educational, Scientific and Cultural Organization (UNESCO): International Standard Classification of Education (ISCED), revisión de 2011.

En el caso a) se confunde no cursar ISCED-2 con no cursar ISCED-3; en el caso b) se confunde cursar ISCED-5 o superior con cursar ISCED-6 o superior.

A veces también se confunde el personal ocupado en Educación con el profesorado (hay trabajadores en el sector de actividad de Educación que no son docentes).

Otra confusión habitual en las investigaciones sobre educación es la de «inmigrante» con «extranjero». Una persona migrante es aquella que marcha de un país a otro por la necesidad de salvaguardar su vida, ya sea mediante la búsqueda de trabajo para poder sobrevivir o escapando de guerras o persecuciones. En el país de donde procede, esa persona será considerada «emigrante» y en el que llega será «inmigrante». Una persona extranjera es aquella que tienen una determinada condición administrativa, que puede variar a lo largo del tiempo. En un país hay personas extranjeras que son migrantes y otras que no los, como por ejemplo, turistas, deportistas de otros países, ejecutivos de multinacionales, miembros del cuerpo diplomático, etc., porque no se han desplazado al país por la necesidad de salvaguardar la vida. Del mismo modo, hay migrantes extranjeros y migrantes con la ciudadanía del país, porque la hayan conseguido administrativamente o porque sean casos de doble nacionalidad. Para evitar estos problemas, algunos estudios (por ejemplo, la encuesta PISA de OCDE) se refiere a: estudiantes nacidos en otro país y a estudiantas cuyos padres han nacido en otro país.

Dos errores habituales en el razonamiento son la denominada falacia de afirmación del consecuente o la confusión entre correlación y causalidad.

Dada una implicación, del tipo «Si A, entonces B», podemos deducir legítimamente que «Si no B, entonces no A», pero nunca «Si B, entonces A». También, si A y B correlacionan, no podemos deducir que A sea causa de B o viceversa (sobre la correlación, véase más adelante).

# **I.- SOBRE LAS TÉCNICAS DE INVESTIGACIÓN CUANTITATIVAS**

---

## **1.- DEFINICIONES PRELIMINARES**

Las técnicas de investigación cuantitativas se basan en el procesamiento de datos numéricos. Podemos obtener los datos de fuentes primarias o secundarias. Los datos primarios son aquellos que recolectamos en nuestra investigación, como, por ejemplo, los resultados de una encuesta. Los datos secundarios son los que se obtienen de fuentes de datos no propias, como, por ejemplo, las estadísticas oficiales.

## **2.- UN EJEMPLO INTRODUCTORIO**

¿Qué efectos producen los sistemas educativos? Las estadísticas oficiales ofrecen una panorámica de los niveles educativos adquiridos por el conjunto de la población. Pero una gran parte de la ciudadanía hace tiempo que abandonó las aulas. Si queremos saber cómo funciona *actualmente* el sistema educativo en un entorno podemos recurrir a una sencilla investigación. Explicándola se intenta mostrar que investigar en educación no siempre requiere diseños complejos o procedimientos costosos.

La educación en España es obligatoria entre los 6-12 años (Educación Primaria) y los 12-16 años (Educación Secundaria). La Educación Secundaria Postobligatoria se divide en bachillerato y Ciclos Formativos de Grado Medio (16-18 años). Después, el estudiante puede acceder a la Educación Superior o Terciaria, que se divide en Universidad o Ciclos Formativos de Grado Superior.

En la asignatura de sociología del primer curso de Magisterio se pregunta a los estudiantes (que tienen unos 18 años): ¿Qué pasó con tus compañeros de primer curso de Educación Primaria (cuando tenían unos 6 años)? La inmensa mayoría de los estudiantes sí que sabe qué trayectoria educativa siguieron sus compañeros y compañeras, con lo que disponemos de una amplia muestra. Hay que precisar que en España hay centros públicos y centros privados de enseñanza. Los segundos recogen 1/3 del estudiantado. Sin embargo, la proporción de estudiantes procedentes de centros privados en la universidad es mayor. Por ello, para evitar un sesgo metodológico, en las respuestas del estudiantado se distinguen los

centros públicos y los privados. Los resultados del último curso de dos grupos se recogen en la tabla 1.

Tabla 1. Resultados de una encuesta: ¿Qué paso con tus compañeros de 1º de Primaria?

	No concluyeron la ESO	Solo cursaron ESO	Solo cursaron bachillerato	Solo cursaron CFGM	Ingresaron en CFGS	Ingresaron en Universidades públicas	Ingresaron en Universidades Privadas	Total
Centros	96	81	64	92	59	122	31	545
%	17,6%	14,9%	11,7%	16,9%	10,8%	22,4%	5,7%	50,1%
Centros	45	45	39	31	44	270	69	543
%	8,3%	8,3%	7,2%	5,7%	8,1%	49,7%	12,7%	49,9%

ESO: Educación Secundaria Obligatoria (12-16 años); CFGM: Ciclos Formativos de Grado Medio (16-18 años); CFGS: Ciclos Formativos de Grado Superior.

Con esta sencilla encuesta hemos conseguido datos sobre una muestra de 1.088 personas de unos 18 años, lo que nos permite una primera aproximación a los problemas del sistema educativo concreto: un porcentaje de estudiantes no concluyen los niveles obligatorios; los que no concluyen más los que solo estudian la Secundaria Obligatoria representan un porcentaje más elevado que el objetivo europeo (que este «abandono educativo temprano», el porcentaje de estudiantado que no cursa Educación Secundaria Postobligatoria, sea inferior al 10%); la mayor parte del estudiantado se cualifica profesionalmente en la universidad y no en la formación profesional (lo que es una disfunción); los centros privados orientan mayoritariamente a la universidad y, en mayor medida, a las universidades privadas, etc. Naturalmente, no podríamos hacer pasar esa pequeña encuesta por un estudio científico riguroso, pero sí que nos sirve, de manera indiciaria, para fijar ulteriores objetos de estudio.

### 3.- ENCUESTAS

Las encuestas son técnicas muy conocidas, por lo que no es preciso explicarlas detalladamente. Simplemente haremos algunas observaciones relacionadas con las encuestas en el ámbito educativo.

#### 3.1.- Sobre la elaboración de cuestionarios

Como norma general no es conveniente obtener datos primarios a partir de los cuestionarios que se apliquen a una muestra que podemos conseguir de fuentes secundarias referidas al universo, porque corremos el riesgo de que nuestras conclusiones respondan a un sesgo de la muestra y sean refutadas por los datos generales.

Es necesario realizar un «pretest» para asegurarnos de que la comprensión de las preguntas del cuestionario es la adecuada y, naturalmente, que sus formulaciones no inducen a una respuesta determinada. Siempre hay que garantizar el anonimato de la persona que responde.

Hay que reducir al máximo las cuestiones a plantear, con el criterio de preguntar solo aquello que ofrecerá informaciones que vamos a procesar (mediante el cálculo de frecuencias, los cruzamientos de datos, etc.).

#### 3.2.- Encuestas y comparaciones

Un error habitual en el diseño de encuestas es confeccionarlas para obtener exclusivamente frecuencias de respuestas. El objetivo de la investigación debe de ser, como dijimos anteriormente, la comparación. Pondremos un ejemplo.

Imaginemos que queremos investigar la autopercepción de la utilidad de la formación académica de las personas que trabajan en educación. Hacemos un cuestionario donde incluimos la pregunta siguiente<sup>1</sup>:

¿En qué grado le sirve su formación académica para el trabajo que realiza? (Utilice una escala del 0 al 10 donde 0: nada, 10: mucho):

0	1	2	3	4	5	6	7	8	9	10

---

<sup>1</sup> Ejemplo tomado de la Encuesta de Calidad de Vida en el Trabajo. Ministerio de Trabajo e Inmigración, 2010.



Después, agrupamos las respuestas 0/1 (muy poca), 2/3 (poca), 4/6 (media), 7/8 (bastante) y 9/10 (mucho) y establecemos la tabla de frecuencias que aparece en la tabla 2.

Tabla 2. Utilidad de la formación académica del personal de educación

Utilidad	Educación
Muy poca	2,6%
Poca	1,7%
Media	17,1%
Bastante	29,2%
Mucha	46,3%
Ns/nc	3,10%

Podemos decir que casi la mitad del personal ocupado en educación considera que es mucha la utilidad de la formación académica o que 3 de cada cuatro personas consideran que la utilidad es bastante o mucha, pero ¿qué significan los datos anteriores? No gran cosa: ¿Son resultados altos? ¿bajos? Estos porcentajes no aportan prácticamente nada al conocimiento del personal ocupado en educación porque no se establece ninguna comparación. Si relacionamos los resultados de la tabla 2 con el resto de secciones de actividad o con los resultados generales (como hemos hecho en la tabla 3) o construimos un índice (como explicaremos más adelante), sí que podemos obtener conclusiones más generales, como, en este ejemplo, que las personas ocupadas en la sección de actividad de educación son las que consideran que su formación es la más útil.

Tabla 3. Utilidad de la formación académica del personal de educación y total

Utilidad	Educación	Total
Muy poca	2,6%	15,1%
Poca	1,7%	6,6%
Media	17,1%	28,3%
Bastante	29,2%	26,3%
Mucha	46,3%	23,7%
Ns/nc	3,10%	6,60%

Si quisiéramos hacer una encuesta sobre este tema (por ejemplo, aplicándola a un centro educativo o a las personas ocupadas en la sección de actividad de educación de un municipio o de una comarca), sería mucho mejor repetir exactamente la

pregunta que ha hecho la encuesta mencionada, para poder comparar los resultados o hacer la encuesta en momentos sucesivos para construir una serie homogénea.

### 3.3.- Atención a «ns/nc»

Aconsejaba Pierre Bourdieu que se prestara atención a los resultados «no sabe/no contesta» de las encuestas. Veamos un ejemplo significativo. En una encuesta del Ministerio de Educación sobre expectativas de nivel máximo de estudios que tienen los alumnos<sup>2</sup> se obtuvieron los resultados que aparecen en la tabla 4. Los datos están desagregados según el nivel formativo de los padres y la edad de los niños que contestan.

Tabla 4. Expectativas del nivel máximo de estudios que tienen alumnos, respecto de su edad y el nivel de estudios de sus padres (porcentajes horizontales)

	Edad	ns/nc	Enseñanza	Bachillerato	Educación
Analfabetos o saben leer y escribir pero no han cursado la Educación Primaria	12 años	47,1%	31,1%	6,8%	15,0%
	14 años	19,9%	35,4%	22,9%	21,8%
	16 años	12,3%	17,0%	37,4%	33,3%
Han cursado la Educación Primaria de manera incompleta o completa	12 años	45,6%	11,3%	7,2%	35,9%
	14 años	12,8%	19,7%	23,2%	44,3%
	16 años	7,6%	6,3%	35,7%	50,4%
Educación Secundaria	12 años	39,9%	5,9%	6,7%	47,5%
	14 años	9,6%	9,5%	18,9%	62,1%
	16 años	5,3%	2,9%	24,6%	67,2%
Educación Superior	12 años	30,5%	4,4%	3,2%	61,9%
	14 años	7,0%	3,1%	6,9%	83,0%
	16 años	5,0%	1,2%	8,5%	85,3%

Véanse los resultados de la columna «ns/nc», no sabe/no contesta. Por ejemplo, los niños de 12 años que son hijos de personas con Educación Superior contestan «ns/nc», en una proporción del 30,5%, pero los hijos de personas analfabetas o que han aprendido a leer y escribir sin asistir a la escuela presentan una proporción de «ns/nc» del 47,1%, un porcentaje casi un 55% superior que en el caso anterior. Todo cuestionario ejerce una cierta violencia porque acota las

<sup>2</sup> MEC (INCE): Sistema estatal de indicadores de la educación 2000, Madrid, MEC, pp. 155-156.

posibilidades de respuesta y suscita connotaciones incontrolables, como parece ser el caso en este ejemplo.

### 3.4.- Un ejemplo de autoevaluación

A continuación se expone un cuestionario de autoevaluación, que adapta el propuesto por Edith Braun y Bettina Hannover, denominado *Das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen* (abreviado BEvaKomp)<sup>3</sup>. Puede ser cumplimentado en diversas ocasiones (por ejemplo, cuando se empiece el estudio en el aula de un tema, en el momento en que esté avanzado o cuando haya finalizado) a fin que la persona pueda comprobar la evolución de su aprendizaje. La puntuación se realiza en una escala de 0 a 5,<sup>4</sup> donde 0 significa máximo desacuerdo con la afirmación y 5 máximo acuerdo.

#### A. Conocimientos

- A.1. Puedo reproducir los conceptos o hechos importantes de esta actividad de aprendizaje.
- A.2. Puedo hacer una síntesis sobre el asunto central de esta actividad de aprendizaje.
- A.3. Puedo exponer de manera clara hechos complicados de esta actividad de aprendizaje.
- A.4. Puedo poner de relieve diferencias y similitudes de los contenidos de esta actividad de aprendizaje con otras actividades de aprendizaje.

#### B. Aptitudes

- B.1. La actividad de aprendizaje me permite enfrentarme mejor a la complejidad de la ciencia actual.
- B.2. Me considero capaz de trasladar algunos elementos de esta actividad de aprendizaje.
- B.3. Con esta actividad de aprendizaje han mejorado mis técnicas de trabajo (búsqueda de bibliografía, documentación, etc.), mis destrezas informáticas o las relacionadas con el conocimiento de otros idiomas.
- B.4. Ha resultado estimulante actividad de aprendizaje para la adquisición de mayores conocimientos.

#### C. Competencias sociales

- C.1. Me veo dispuesto/a a elaborar autónomamente una exposición pública del contenido de esta actividad de aprendizaje, ampliándolo con asuntos y ejemplos de otros ámbitos.

---

<sup>3</sup> Véase: Braun, Edith (2007): *Das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen - BEvaKomp*. Göttingen: Vandenhoeck & Ruprecht unipress; Braun, Edith; Hannover, Bettina (2008): «Kompetenzmessung und Evaluation von Studienerfolg», en: Jude, Nina; Hartig, Johannes; Klieme, Eckhard (eds.): *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden*. Berlín: Bundesministerium für Bildung und Forschung (BMBF); Jude, Nina; Hartig, Johann; Klieme, Eckhard (eds.) (2008): *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden*. Berlín: Bundesministerium für Bildung und Forschung. (disponible en [www.bmbf.de](http://www.bmbf.de))

<sup>4</sup> La escala 0 a 5 tiene 6 opciones; por ello no hay un valor central, lo que siempre es recomendable a la hora de diseñar encuestas.

C.2. He tratado del contenido de esta actividad de aprendizaje con personas de mi entorno, lo que ha resultado positivo.

C.3. Esta actividad de aprendizaje aporta información relevante sobre mis relaciones con la sociedad en general.

C.4. Resulta relevante el contenido de esta actividad de aprendizaje para mis proyectos de estudio o trabajo.

## 4. ÍNDICES Y FÓRMULAS

### 4.1.- Índices

Un índice es un valor numérico (aritmético) que pretende dar cuenta de un fenómeno. En general pretendemos simplificar un conjunto de datos mediante un dato. Pondremos un ejemplo.

Tenemos dos centros educativos de Educación Secundaria postobligatoria, con diversas especialidades de bachillerato y formación profesional, y queremos establecer un índice para evaluar el sexismo en la distribución del estudiantado. Los datos iniciales se encuentran en la tabla 5.

Tabla 5. Estudiantes de especialidades en dos centros educativos<sup>5</sup>

Centro educativo	Mujeres	Hombres
<i>Centro educativo 1</i>		
Especialidad A	42	21
Especialidad B	40	24
Especialidad C	8	20
<i>Total estudiantes centro educativo 1</i>	90	65
<i>Centro educativo 2</i>		
Especialidad D	14	36
Especialidad E	30	25
Especialidad F	43	59
Especialidad G	4	21
Especialidad H	19	12
Especialidad I	25	29
Especialidad J	1	32
Especialidad K	32	7
Especialidad L	12	16
Especialidad M	51	41
<i>Total estudiantes centro educativo 2</i>	231	278

A simple vista es muy difícil apreciar qué especialidad presenta una distribución más sexista. Por ello, definiremos un índice ( $I_e$ ) del modo siguiente.

---

<sup>5</sup> Estos datos están tomados de un estudio sobre la localidad de Manises (Comunidad Valenciana).

-Consideraremos el porcentaje de mujeres en cada especialidad ( $M_e$ ), con la fórmula siguiente:  $M_e/(M_e + H_e)$ , es decir, número de mujeres de cada especialidad dividido por el número de mujeres y hombres de la misma especialidad (daría igual hacerlo con la proporción de hombres).

-A continuación, consideraremos el porcentaje de mujeres en cada centro educativo ( $M_c$ ), con la fórmula siguiente:  $M_c/(M_c + H_c)$ , es decir, número de mujeres de cada centro dividido por el número de mujeres y hombres del mismo centro.

-Calcularemos la diferencia entre la proporción de mujeres en cada especialidad respecto de la proporción de mujeres en cada centro. Calcularemos el valor absoluto de esta diferencia (por el procedimiento de elevar la cantidad al cuadrado y calcular su raíz cuadrada). [Como se sabe, el valor absoluto de  $+x$  y  $-x$ , que se formulan así:  $|x|$  y  $|-x|$ , son  $x$ ; en ambos casos, si los elevamos al cuadrado  $x$  o  $-x$  y hacemos su raíz cuadrada el resultado es igualmente  $x$ ].

-Restaremos la cantidad obtenida (un valor entre 0 y 1) de la unidad  $y$ , para maximizar la diferencia entre las diversas especialidades, la elevaremos al cubo. Por último, la multiplicaremos por 10 para que el valor resultante esté en el rango 0-10 [la multiplicaríamos por 100 si quisieramos trabajar con el rango 0-100].

Todo este proceso se puede resumir en la fórmula siguiente:

$$I_e = 10 \left[ 1 - \left| \left( \frac{M_e}{M_e + H_e} - \frac{M_c}{M_c + H_c} \right) \right| \right]^3$$

O, sustituyendo el valor absoluto de la diferencia por la raíz del cuadrado, tenemos:

$$I_e = 10 \left[ 1 - \sqrt{\left( \frac{M_e}{M_e + H_e} - \frac{M_c}{M_c + H_c} \right)^2} \right]^3$$

A partir de esta fórmula, se obtienen los resultados que aparecen en la tabla 6:

Tabla 6. Índice de sexismo en la distribución de estudiantado por especialidades

	Índice
Centro educativo 1	
Especialidad A	7,6
Especialidad B	8,7
Especialidad C	3,5

Centro educativo 2	
Especialidad D	5,6
Especialidad E	7,5
Especialidad F	9,1
Especialidad G	3,5
Especialidad H	5,9
Especialidad I	9,7
Especialidad J	1,9
Especialidad K	2,5
Especialidad L	9,3
Especialidad M	7,3

El índice nos permite visualizar fácilmente que la distribución más sexista se da en las especialidades J y K del centro 2, y las menos sexistas se encuentran en las especialidades I y L del mismo centro.

#### 4.2.- Fórmulas

Además de la elaboración de índices, también podemos expresar de manera simplificada un conjunto de datos o una relación mediante una fórmula, generalmente una ecuación. Pondremos un ejemplo.

Hemos analizado concienzudamente las bases de datos de matriculación de una universidad grande (con unos 50.000 estudiantes cada curso). De los datos de los cursos 2009/10-2014/15<sup>6</sup>, seleccionamos los estudiantes que cambiaron de grado (esto es, se reubicaron en otra carrera) y realizamos un recuento de los que se cambiaron el primer año de su matrícula, el segundo, etc. ¿Podemos establecer una pauta? Los resultados se muestran en la tabla 7

Tabla 7. Reubicaciones según el año de cambio (2009/10-2014/15)

Año	Reubicaciones
1r año	1789
2º año	371
3r año	177
4º año	122

<sup>6</sup> Universidad de València. En el hemisferio Norte, los cursos comienzan después del verano (septiembre) y concluyen antes de las vacaciones de verano (junio o julio), por lo que se aluden con dos años naturales.

Total	2459
-------	------

Para buscar una fórmula, calcularemos la proporción que supone cada año en el conjunto de reubicaciones, representaremos estas proporciones en un gráfico y calcularemos la línea de tendencia (potencial) y su ecuación (más adelante se explicarán ambas cosas). La ecuación resultante es:  $y = 0,6717x^{-1,967}$  en la que «x» es el curso desde la primera matrícula e «y» la proporción de estudiantes que se reubicará en ese curso respecto de total de reubicados. Ahora bien, la fórmula anterior ( $y = 0,6717x^{-1,967}$ ) no es sencilla de recordar. Pero podemos enunciar una aproximación más simple y fácil de manejar:

$$y = \frac{2}{3} x^{-2}$$

## 5. RELACIÓN DE DOS VARIABLES

### 5.1. Determinación de tendencias

Ante una serie de datos podemos establecer tendencias de un modo razonado. Pondremos un ejemplo. El Instituto Nacional de Estadística (INE) realiza trimestralmente una Encuesta de Presupuestos Familiares (EPF). Con esta encuesta se establecen los porcentajes que las familias destinan a cada grupo de gasto. En la tabla 8 tenemos los datos, desagregados por niveles de ingresos mensuales regulares y netos de los hogares.

Tabla 8. Porcentaje de gasto en cada grupo de gasto del presupuesto familiar (porcentajes horizontales)

Grupos	G01	G02	G03	G04	G05	G06	G07	G08	G09	G10	G11	G12
Total	15,04	1,91	5,10	31,69	4,25	3,56	11,52	2,89	5,83	1,38	9,31	7,52
...- 499 €	20,38	2,56	3,40	42,34	2,94	2,59	8,43	3,14	3,22	0,52	4,96	5,53
500-999 €	18,29	1,90	3,91	41,77	3,69	3,73	7,37	3,13	3,76	0,49	5,19	6,77
1.000-1.499 €	16,94	2,23	4,67	35,07	3,72	3,59	10,50	3,20	4,76	0,65	7,45	7,21
1.500-1.999 €	15,46	2,03	5,10	31,13	4,04	3,83	11,54	3,03	5,61	0,93	9,30	8,00
2.000-2.499 €	14,73	2,02	5,30	29,87	4,24	3,77	12,27	3,01	6,13	1,33	9,82	7,52
2.500-2.999 €	14,09	1,84	5,64	28,26	4,00	3,47	13,44	2,78	6,69	1,55	10,60	7,63



3.000-4.999 €	12,86	1,59	5,69	27,09	4,78	3,49	13,06	2,50	7,28	2,32	11,65	7,70
5.000-... €	10,33	1,25	5,58	27,91	6,82	2,66	12,21	2,20	6,97	3,42	12,33	8,33

Grupo 1. Alimentos y bebidas no alcohólicas. Grupo 2. Bebidas alcohólicas, tabaco y narcóticos. Grupo 3. Artículos de vestir y calzado. Grupo 4. Vivienda, agua, electricidad, gas y otros combustibles. Grupo 5. Mobiliario, equipamiento del hogar y gastos corrientes de conservación de la vivienda. Grupo 6. Salud. Grupo 7. Transportes. Grupo 8. Comunicaciones. Grupo 9. Ocio, espectáculos y cultura. Grupo 10. Enseñanza. Grupo 11. Hoteles, cafés y restaurantes. Grupo 12. Otros bienes y servicios. Fuente: INE: Encuesta de Presupuestos Familiares, 2015.

Observemos el grupo G04 (vivienda), el que representa un porcentaje mayor, y el grupo G01 (alimentos). Podemos observar que conforme aumentan los ingresos mensuales disminuye el porcentaje que esos grupos representan en los presupuestos familiares. Otros grupos, como el G05 (mobiliario) o el G07 (transportes), incrementan el porcentaje conforme aumentan los ingresos. Podemos resumir esta apreciación formulando un enunciado general: en el caso de los capítulos de gasto correspondientes a necesidades básicas, el porcentaje disminuye conformen aumentan los beneficios; en el caso de capítulos de gasto relativos a bienes de distinción, sucede lo contrario.

Establecida esta tendencia general, podemos plantearnos qué sucede con otros grupos. Por ejemplo, el G02 (alcohol, tabaco) se comporta como el G01 (alimentos). El G06 (salud) presenta aumentos y descensos, ya que los gastos en salud presentan la doble característica: son una necesidad básica y pueden ser un bien de distinción (p. ej., intervenciones estéticas). Obsérvese también como el G10 (educación) presenta la configuración característica de un bien de distinción.

De lo que se trata no es de ver qué grupos incrementan su porcentaje o lo disminuyen según aumenten o no los niveles de ingresos, sino de establecer una pauta general. Para intentar dar un rigor mayor a este objetivo, que en el ejemplo se puede realizar «a simple vista», podemos recurrir al trazo de líneas de tendencia.

## 5.2. Líneas de tendencia

Veamos otro ejemplo. En la tabla 9 hay diversos países de la Unión Europea y los datos que corresponden a dos variables: la inversión pública en educación (como porcentaje del Producto Nacional Bruto) y el porcentaje de estudiantado de 15

años con bajo rendimiento en matemáticas (según las pruebas PISA). En ambos casos, los datos están referidos al año 2012<sup>7</sup>.

Tabla 9. Inversión pública en educación y bajo rendimiento en matemáticas en países de la Unión Europea

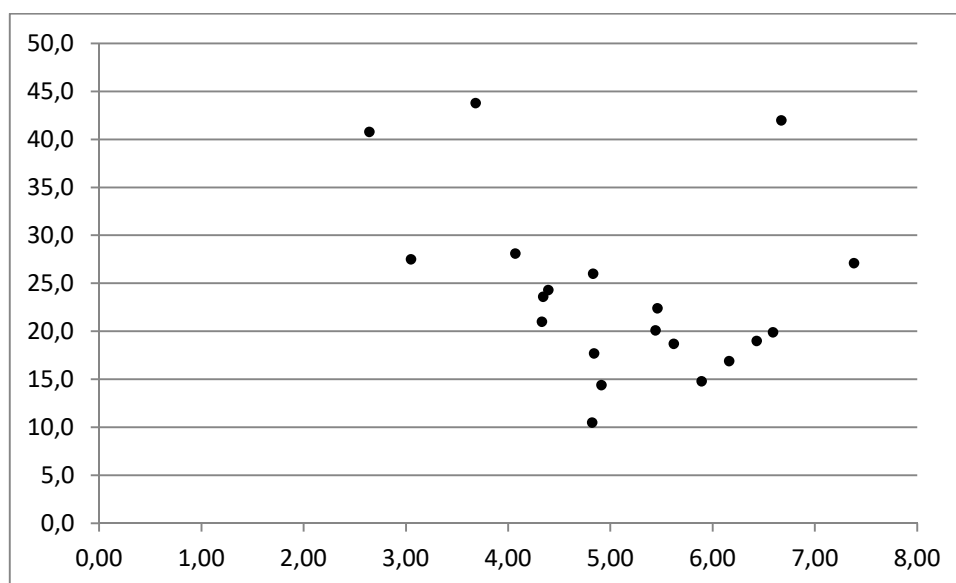
País	Inversión pública	Bajo rendimiento
Alemania	4,84	17,7
Austria	5,62	18,7
Bélgica	6,43	19,0
Bulgaria	3,68	43,8
Chequia	4,33	21,0
Chipre	6,67	42,0
Eslovaquia	3,05	27,5
Eslovenia	5,44	20,1
España	4,34	23,6
Estonia	4,82	10,5
Francia	5,46	22,4
Holanda	5,89	14,8
Hungría	4,07	28,1
Irlanda	6,16	16,9
Letonia	6,59	19,9
Lituania	4,83	26,0
Luxemburgo	4,39	24,3
Polonia	4,91	14,4
Rumanía	2,64	40,8
Suecia	7,38	27,1

Vamos a trasladar esta tabla a una hoja de cálculo (utilizaremos el programa Excel, aunque daremos instrucciones también sobre otros programas).

Una vez que hemos trasladado la tabla anterior, seleccionamos los datos (manteniendo pulsado el botón izquierdo y pasando el cursor por encima de las celdas) y hacemos clic en la pestaña <Insertar> y después en <Gráficos> seleccionamos <Insertar gráfico de dispersión XY o de burbujas>. Entonces aparece un gráfico, como este:

<sup>7</sup> Fuente: Eurostat, códigos: educ\_uoe\_fine06 y sdg\_04\_40

Gráfico 1.- Distribución de valores en el espacio cartesiano



¿Cómo podemos apreciar la relación entre las dos variables la inversión pública (que aparece en el eje X) y el bajo rendimiento (que aparece en el eje Y)?

Podemos recurrir, en primer lugar, a trazar una línea de tendencia.

Como se sabe, en un plano cartesiano (donde están definidos los ejes X e Y), toda línea (recta o curva) se puede expresar con una ecuación, y viceversa: toda ecuación define una línea. Los tipos de ecuaciones que definen líneas pueden ser:

Ecuación de la línea recta  $y = ax+b$  (en algunos manuales:  $y = mx + n$ )

Ecuación exponencial  $y = ae^{bx}$

Ecuación logarítmica  $y = a\ln(x)+b$

[utilizando logaritmos naturales o neperianos, que recuerdese que tienen como base el número e, de manera que  $\ln(e^x)=x$ ]

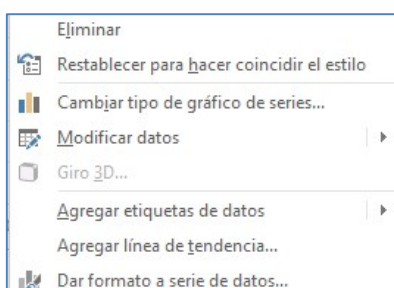
Ecuación polinómica (orden o grado 2) :  $y = ax^2+bx+c$

(orden o grado 3)  $y = ax^3+bx^2+cx+d$

(orden o grado 5)  $y = ax^4+bx^3+cx^2+dx+e$ , etc.

Ecuación potencial  $y = ax^b$

Para trazar la línea de tendencia ponemos el cursor sobre uno de los puntos del gráfico y hacemos clic en el botón derecho del ratón. Se despliega este menú (imagen 1):

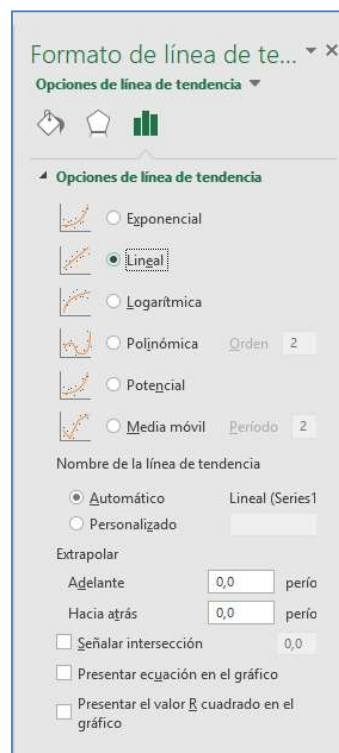


[Imagen 1]

Hacemos clic en

<Agregar línea de tendencia...>

y aparecerá el siguiente menú (imagen 2):



[Imagen 2]

Marcamos una línea de tendencia y también <Presentar ecuación en el gráfico> Entonces el programa elabora los gráficos. Veamos los diversos casos (salvo el último: media móvil).

Gráfico 2.- Línea de tendencia lineal (recta)

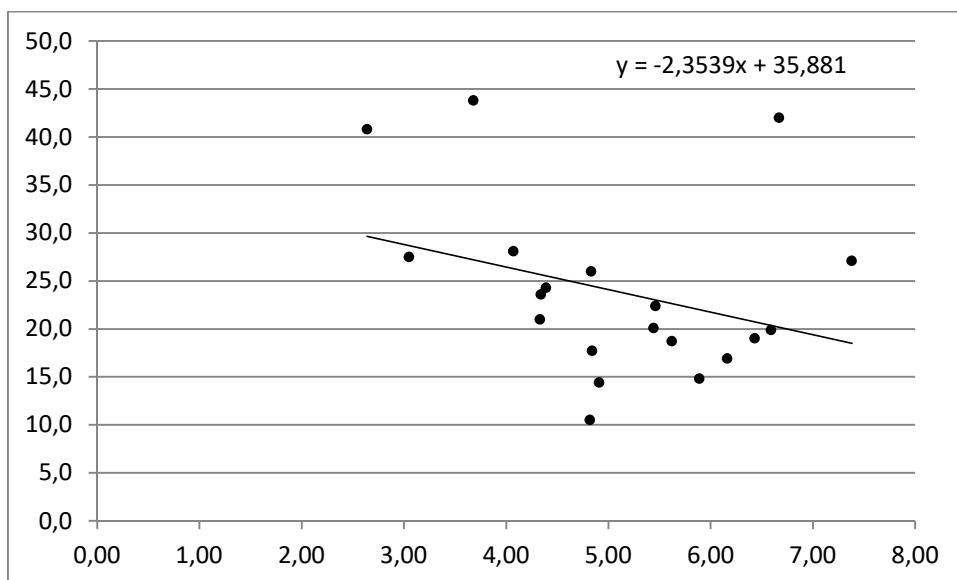


Gráfico 3.- Línea de tendencia exponencial

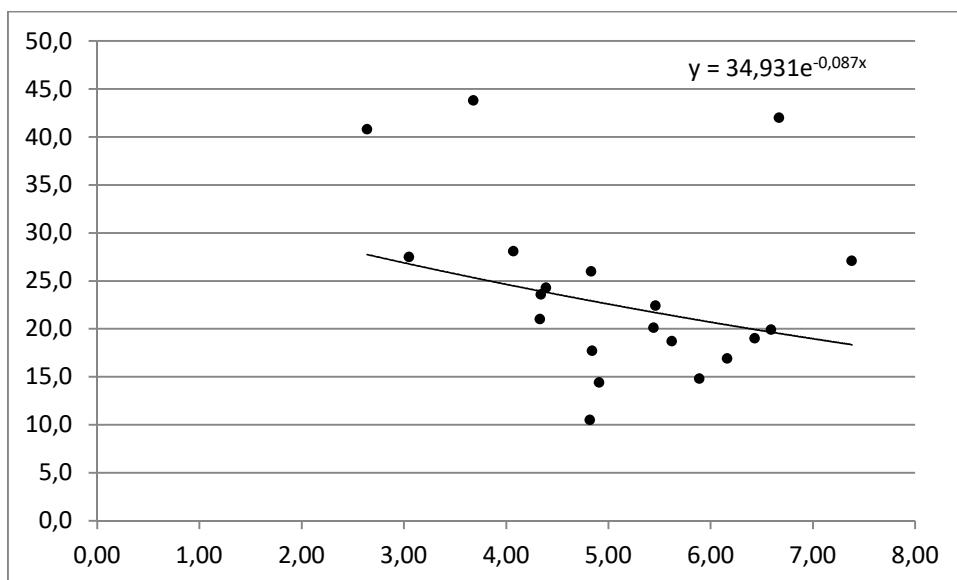


Gráfico 4.- Línea de tendencia logarítmica

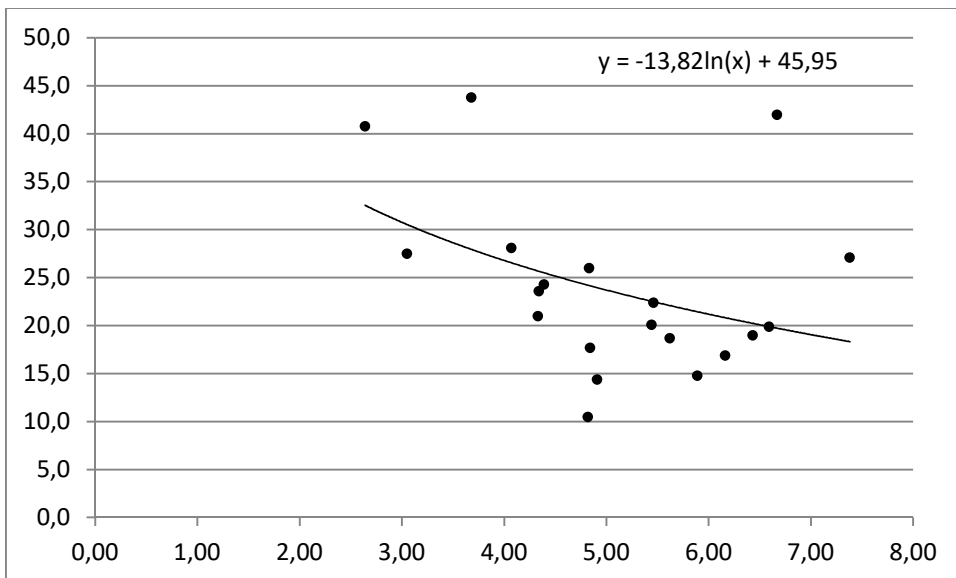


Gráfico 5.- Línea de tendencia polinómica (orden 2)

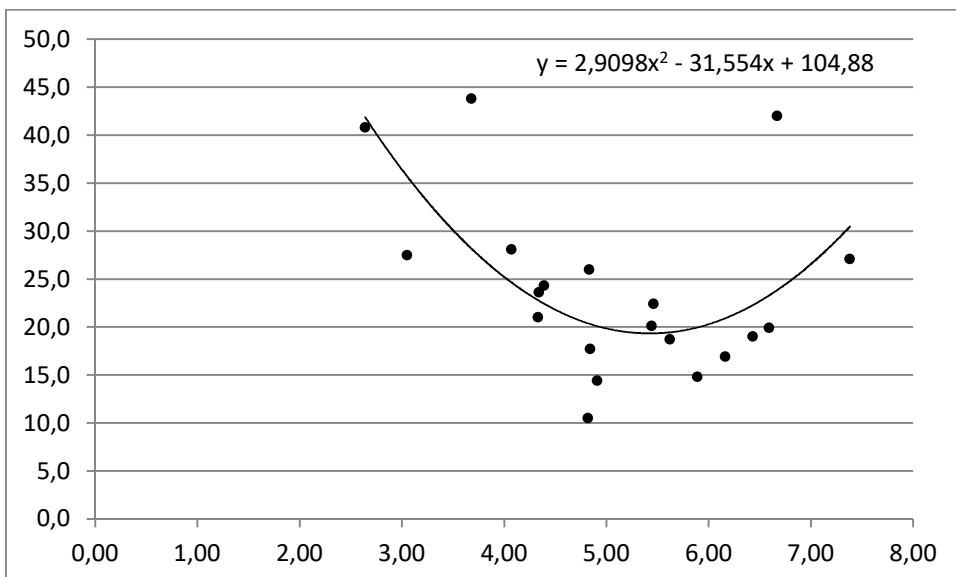


Gráfico 6.- Línea de tendencia polinómica (orden 3)

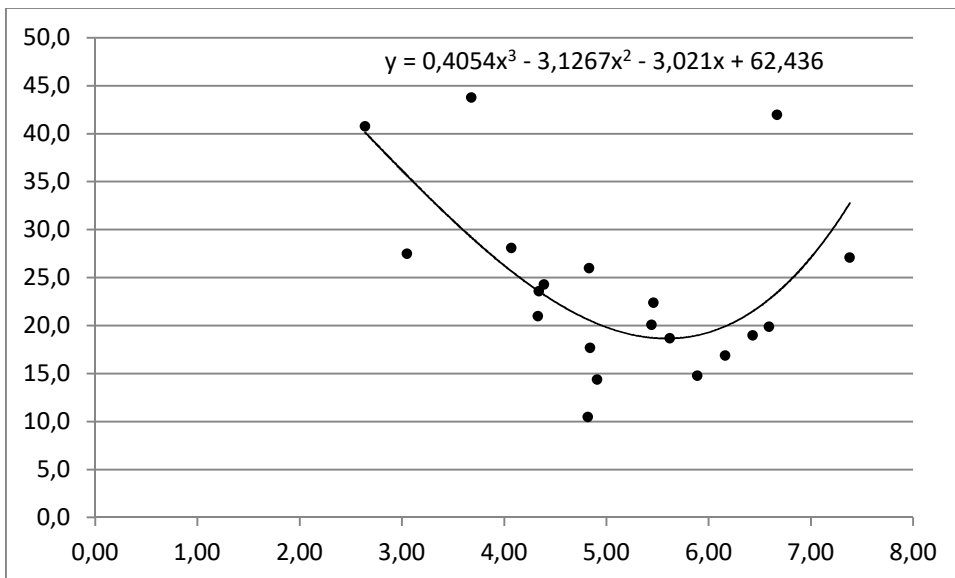


Gráfico 7.- Línea de tendencia polinómica (orden 4)

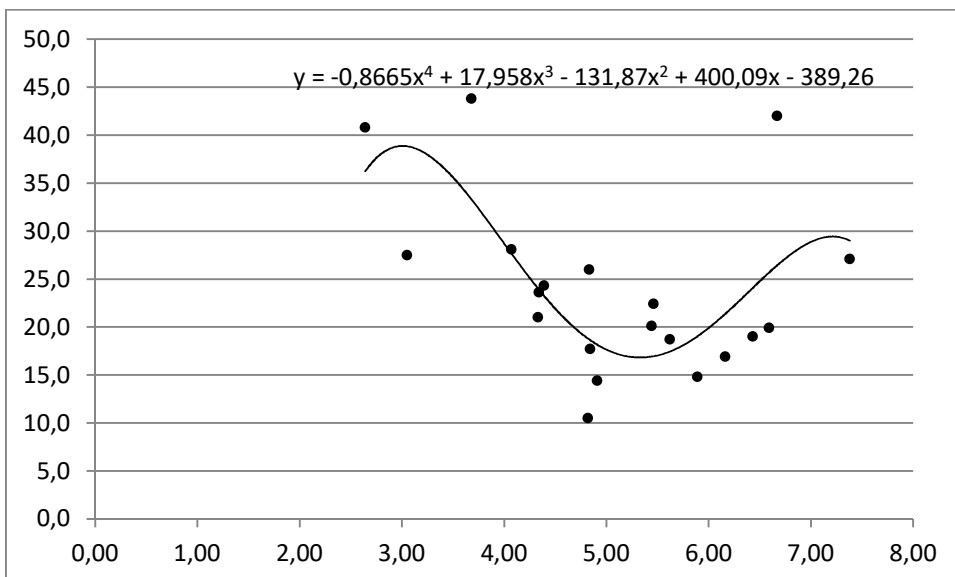


Gráfico 8.- Línea de tendencia polinómica (orden 5)

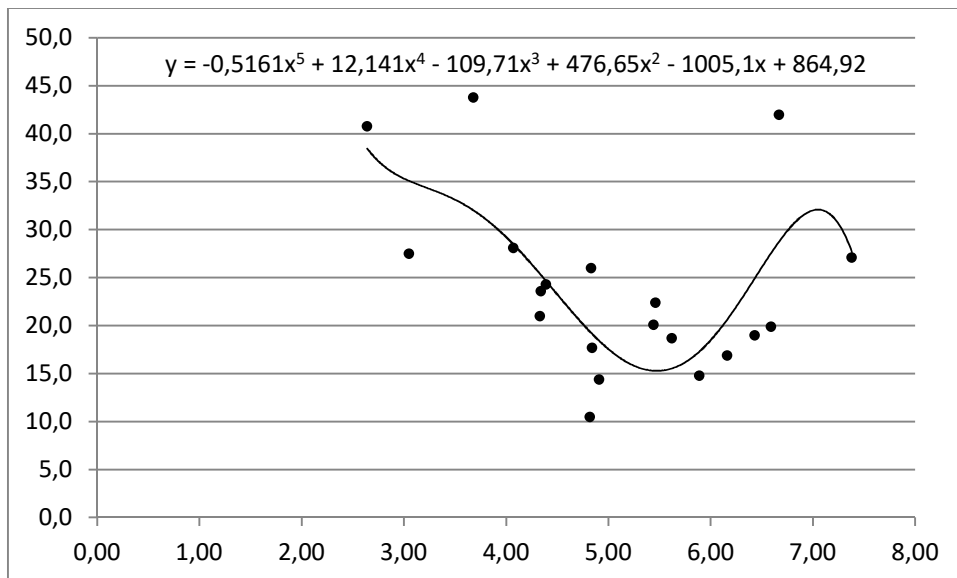


Gráfico 9.- Línea de tendencia polinómica (orden 4)

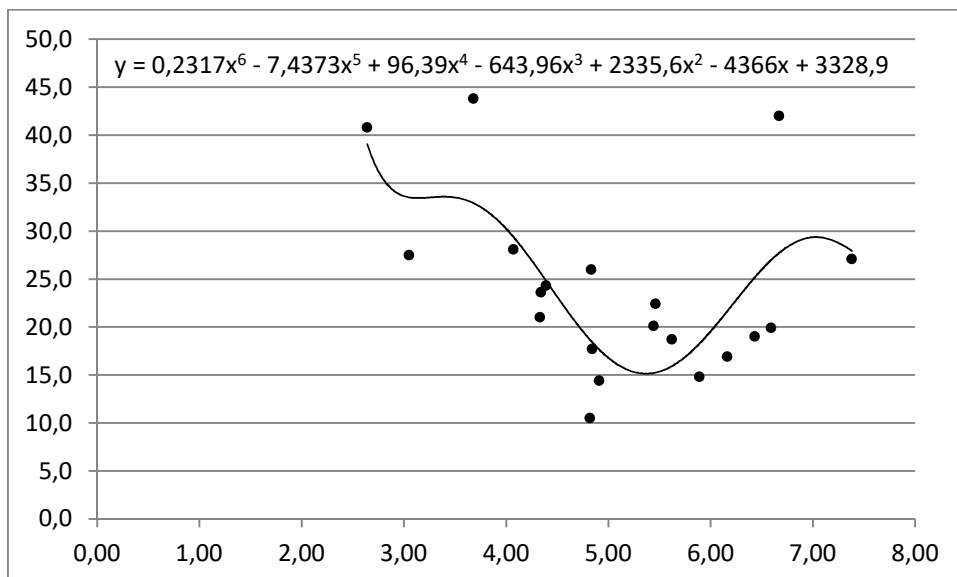
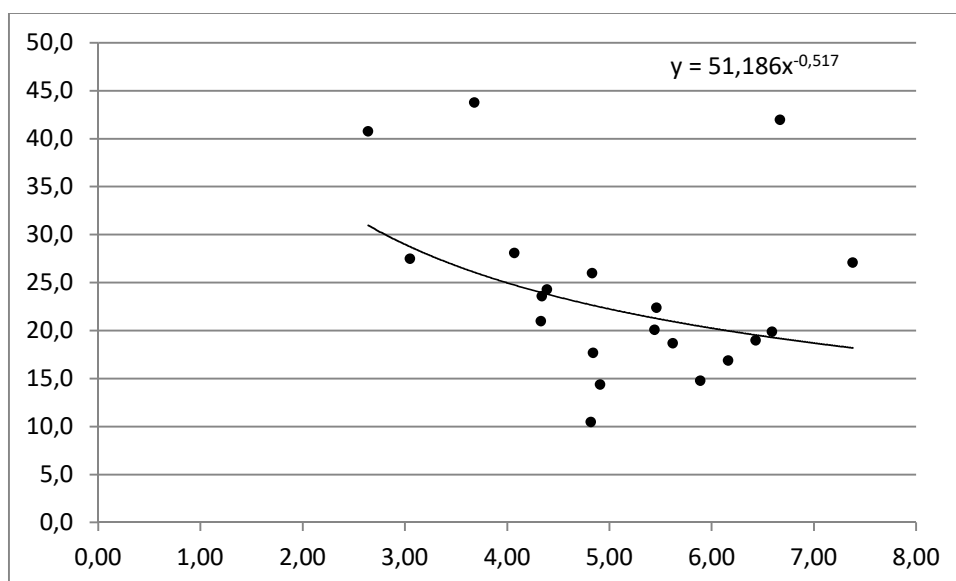




Gráfico 10.- Línea de tendencia potencial



Después de este ejercicio gráfico podemos preguntarnos: ¿qué es realmente una línea de tendencia?

Observemos el primer caso, el más simple, la línea de tendencia recta. Podríamos decir que, de las infinitas rectas del plano cartesiano, la línea de tendencia sería aquella que definiría distancias menores a los puntos de la distribución.

La línea de tendencia y la ecuación asociada a ella nos permitirían hacer una estimación del tipo:

¿Qué valor de Y correspondería al valor de un determinado valor de X?

O al revés:

¿Qué valor de X correspondería al valor de un determinado valor de Y?

Lo podríamos hacer gráficamente, trazando las abscisas y ordenadas hasta la línea de tendencia o aritméticamente, utilizando la ecuación de la línea (y asignando un valor de X o de Y y calculando respectivamente Y o X).

Pero todavía no haremos estos cálculos, porque necesitamos antes depurar los datos, y para ello podemos utilizar el coeficiente de correlación, cuyo cálculo explicaremos a continuación. Cuando depuremos los datos podremos además plantearnos una cuestión decisiva: ¿qué línea de tendencia hemos de elegir?

### 5.3. Coeficiente de correlación

Un procedimiento más sofisticado que la línea de tendencia para establecer la relación entre dos variables es calcular su coeficiente de correlación, aquí utilizaremos el coeficiente de Pearson-Bravais (para simplificar, de Pearson).

El coeficiente de correlación o coeficiente de Pearson se suele expresar mediante la letra  $r$ , la mayúscula,  $R$ , o la letra griega equivalente  $\rho$ ; aquí utilizaremos  $R$ , que es como aparece en el programa Excel.

El coeficiente de Pearson expresa la relación entre dos variables,  $X$  e  $Y$ . El valor de  $R$  puede oscilar entre  $-1$  y  $1$ . En caso de que haya una correlación intensa y directa entre  $X$  e  $Y$ , el valor de  $R$  se aproxima a  $1$ . Esto quiere decir que cuando los valores de una variable crecen, los correspondientes de la otra suelen crecer al mismo ritmo, y viceversa. En caso de que no exista prácticamente correlación, el valor de  $R$  se aproxima a  $0$ . Entonces, la oscilación de una variable no se corresponde con la de la otra. En caso de que una variable aumente y la otra disminuya al mismo ritmo, el valor de  $R$  se acerca a  $-1$ . Hablaremos entonces de una correlación intensa, pero inversa.

Correlación inversa				Ausencia de correlación				Correlación		
-1,0	-0,8	-0,6	-0,4	-0,2	0,0	0,2	0,4	0,6	0,8	1,0

En el caso de una correlación inversa, simplemente con emplear la variable inversa de una de las dos (digamos  $X^{-1}$  o  $Y^{-1}$ ),  $R$  cambia su signo. Por ello, muchas veces podemos utilizar el valor absoluto de  $R$  (es decir:  $|R|$ ), sin tener en cuenta si es positivo o negativo su valor (por esta razón, en algunos manuales de estadística se dice que  $R$  oscila entre  $0$  y  $1$ ). Por tanto:

$$-1 \leq R \leq 1 \quad ; \quad 0 \leq |R| \leq 1$$

El coeficiente de correlación de Pearson se define como el cociente de la covarianza de dos variables,  $X$  e  $Y$ , por el producto de sus desviaciones típicas.

$$R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Veamos estos conceptos.

La covarianza,  $\sigma_{xy}$ , se define como el sumatorio del producto de la diferencia de cada valor de  $x$  ( $x_i$ ) menos su media ( $\bar{x}$ ) por el correspondiente valor de  $y$  ( $y_i$ ) menos su media ( $\bar{y}$ ), dividido por el número de valores  $-1$ .

$$\sigma_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

La desviación típica, que se expresa con la letra  $s$  o la letra griega equivalente sigma,  $\sigma$ , equivale a la raíz cuadrada de la varianza. La varianza es la suma de la diferencia de cada valor con la media dividida por el número de valores.

$$\sigma_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

De manera manual, se puede calcular el coeficiente de correlación de Pearson mediante la fórmula siguiente:

$$R_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Dados  $n$  valores de las variables  $X$  e  $Y$ , para calcular manualmente un coeficiente de correlación de este modo es conveniente hacer una tabla del tipo siguiente y averiguar los datos sombreados (recuérdese que el signo  $\Sigma$  significa sumatorio).

X	Y	XY	X <sup>2</sup>	Y <sup>2</sup>
x <sub>1</sub>	y <sub>1</sub>	x <sub>1</sub> y <sub>1</sub>	x <sub>1</sub> <sup>2</sup>	y <sub>1</sub> <sup>2</sup>
x <sub>2</sub>	y <sub>2</sub>	x <sub>2</sub> y <sub>2</sub>	x <sub>2</sub> <sup>2</sup>	y <sub>2</sub> <sup>2</sup>
...	...	...	...	...
x <sub>n</sub>	y <sub>n</sub>	x <sub>n</sub> y <sub>n</sub>	x <sub>n</sub> <sup>2</sup>	y <sub>n</sub> <sup>2</sup>
$\Sigma x_n$	$\Sigma y_n$	$\Sigma x_n y_n$	$\Sigma x_n^2$	$\Sigma y_n^2$
$\Sigma x_n^2$	$\Sigma y_n^2$			

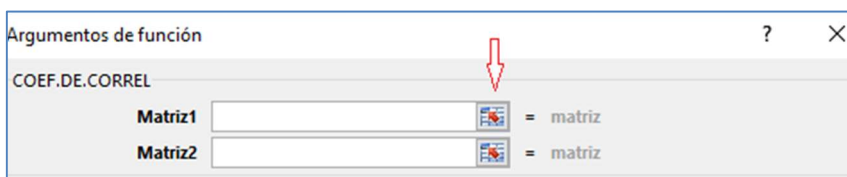
Afortunadamente, algunos programas informáticos habituales como las hojas de cálculo (Excel, etc.) permiten el cálculo inmediato del coeficiente de correlación de Pearson. Para ello procederemos del modo siguiente.

Marcamos una celda vacía y hacemos clic en el botón ***fx*** que hay a la izquierda de la barra de fórmulas. En el menú desplegable, o bien escribimos en la ventana superior <COEF.CORREL> o <PEARSON> o bien seleccionamos en la segunda ventana la categoría <Estadísticas> y seleccionamos después <COEF.CORREL> o <PEARSON>. Entonces aparece la ventana:



[Imagen 3]

O bien ponemos en la ventana el valor de la matriz (las celdas que ocupa cada variable. Por ejemplo: B2: B21) o (recomendable) hacemos clic en el botón de la matriz y luego marcamos, presionando el botón izquierdo, las celdas de cada variable.



[Imagen 4]

Al concluir la segunda matriz, se hace clic en <Aceptar> y aparecerá la correlación en la celda seleccionada.

También cuando se traza la línea de tendencia, el programa da la opción de proporcionar el dato de  $R^2$  (véase la imagen 2: <Presentar el valor  $R^2$  en el gráfico>). Hay que tener en cuenta que el mismo  $R^2$  puede corresponder a dos valores de  $R$ :  $+R$  y  $-R$ . ¿Cómo sabemos si el valor de  $R$  es positivo o negativo? Si la línea es creciente: el valor de  $R$  es positivo; si la línea es decreciente (como en el ejemplo anterior), el valor de  $R$  es negativo.

En el ejemplo anterior,  $R = -0,319$  (a veces se expresa porcentualmente como  $-31,9\%$ )

## 5.4. Interpretación de R y de R<sup>2</sup>

Resulta trivial afirmar que  $R_{xy} = R_{yx}$ . Da igual que variable ubiquemos en el eje X o que variable ubiquemos en el eje Y (o, según la imagen 3, qué variable atribuyamos a la matriz 1 o a la matriz 2). En ambos casos, el valor de R será el mismo.

En la bibliografía estadística hay tablas con criterios para determinar si un valor de R es bajo, medio, alto, etc. Estas apreciaciones son convencionales. Hay que tener en cuenta que no es lo mismo establecer una correlación en ciencias naturales o de la salud que en ciencias humanas o sociales, en las que un valor de 0,6 o 0,7 ya es muy elevado. Además, como hemos dicho, el valor de R es muy sensible a puntuaciones extremas (que a veces responden a causas espúreas). También hay que considerar el número de casos a la hora de apreciar el valor de un coeficiente de correlación. No es igual obtener una correlación de 0,5 con 20 que con 100 casos. En este segundo caso, el coeficiente será más relevante.

Convencionalmente se afirma a veces que R<sup>2</sup> corresponde al porcentaje de una variable explicada por la otra, pero esta afirmación exigiría indagaciones ulteriores.

Es muy importante advertir que *correlación* no es lo mismo que *causalidad*. Lo que el coeficiente expresa es la relación mutua de dos variables, pero no establece que, cuando el coeficiente es elevado, una sea la causa de la otra. Desgraciadamente son muy frecuentes las investigaciones que concluyen afirmando una relación de causalidad a partir del establecimiento de un coeficiente elevado. A veces un coeficiente de correlación elevado se produce porque las dos variables se relacionan con una tercera. Por utilizar un símil: dos bailarines pueden ir totalmente acompasados, pero no porque el movimiento de uno sea la *causa* del movimiento del otro, sino porque los dos armonizan su coreografía con la música que suena.

## 5.5. Depuración de datos y elección de la línea de tendencia

Como hemos dicho, R es muy sensible a puntuaciones extremas. Si de la relación de países eliminamos Chipre, R alcanza el valor -0,545; si eliminamos Suecia alcanza -0,690; y si eliminamos a Estonia alcanzamos un valor de -0,743, que es un coeficiente de correlación muy importante. ¿Qué significa esto? Hay una tendencia general clara (con una correlación de -0,743), pero hay tres países discordantes por algún motivo (que no tiene necesariamente que ser el mismo en todos los casos). Para localizar estos casos discordantes resulta útil visualizar qué valores se alejan de las líneas de tendencia.

Si no consideramos estos países discrepantes, cambian las ecuaciones de las líneas de tendencia. Por ejemplo, la de la línea de tendencia recta pasa a ser:

$$y = -5,3645x + 49,558$$

O la línea de tendencia exponencial:

$$y = 95,702x^{-0,936}$$

Por lo tanto, ya podemos proceder a plantearnos las cuestiones anteriores: ¿Qué valor de Y correspondería al valor de un determinado valor de X? O al revés: ¿Qué valor de X correspondería al valor de un determinado valor de Y?

Por ejemplo, si utilizamos la ecuación de la línea de tendencia exponencial podemos observar como con una inversión pública del 5% del PIB (valor X), el bajo rendimiento en matemáticas se tiene que colocar en el 21,21% (valor Y) y que con un punto más de inversión, desciende al 17,88%.

Pero ¿qué tipo de línea de tendencia debemos escoger?

Muchas veces se usa la línea de tendencia recta, pero no parece la más conveniente en algunos casos.

En el caso del ejemplo anterior o en otros relacionados con dinámicas de saturación, es recomendable el uso de líneas de tendencia logarítmicas o exponenciales. Una dinámica de saturación es aquella en la que el incremento de una variable produce incrementos en la otra, pero cada vez a un menor ritmo. El ejemplo clásico es el resultado de añadir azúcar al café. La primera cucharada supone un incremento de dulzor en la bebida. La segunda, un incremento menor, y así sucesivamente. Esas dinámicas se pueden expresar con curvas logarítmicas. En el ejemplo anterior, incrementos de inversión educativa correlacionaban con disminuciones del bajo rendimiento en matemáticas. Es por eso que podemos utilizar una línea de tendencia exponencial o calcular el inverso de la variable rendimiento y aplicar una línea de tendencia logarítmica. En los gráficos 11 y 12 se representan estas líneas con los datos depurados de la tabla 9. Obsérvese que las líneas son prácticamente simétricas.

Gráfico 11. Datos de la tabla 9 (depurados) con línea de tendencia exponencial

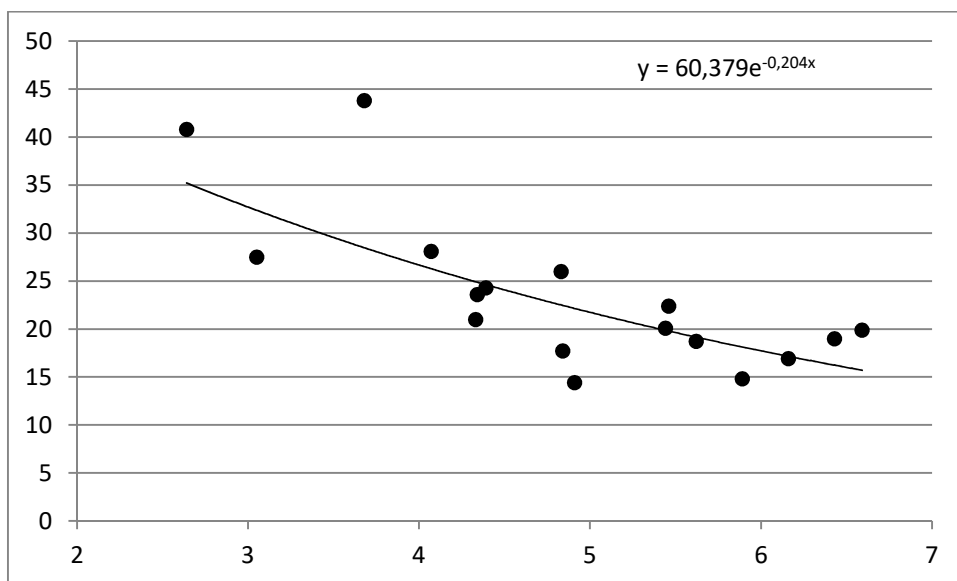
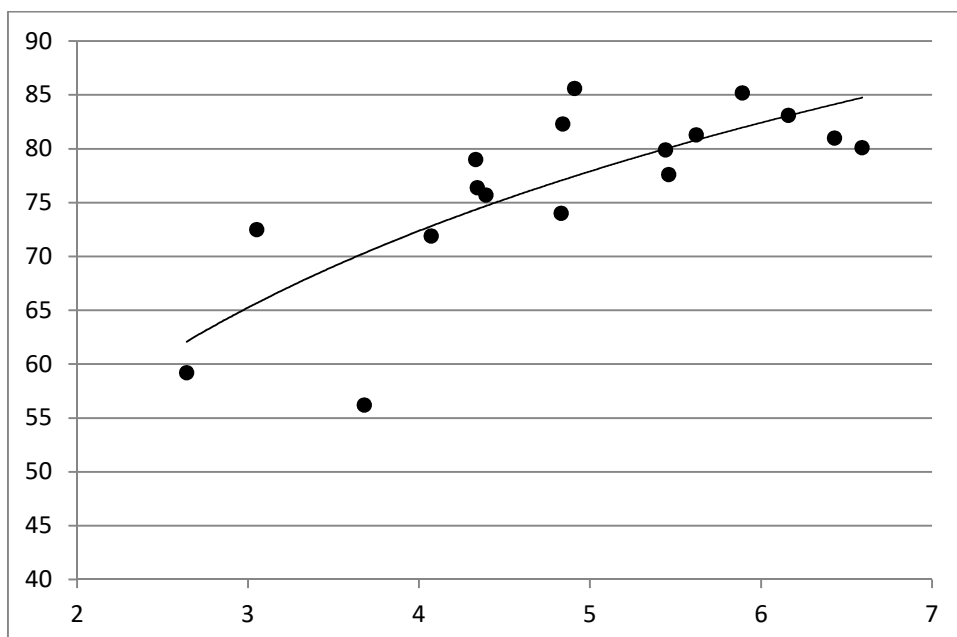


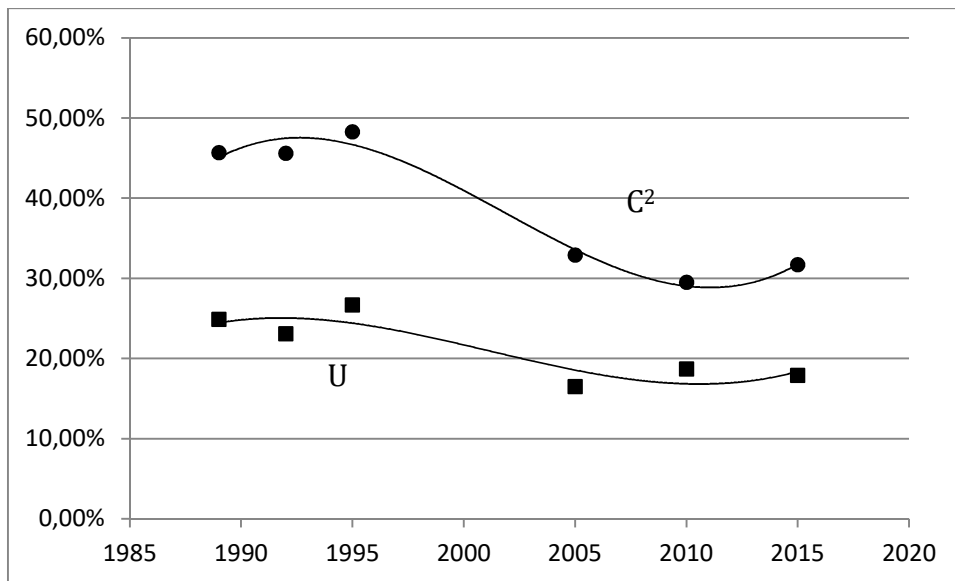
Gráfico 12. Datos de la tabla 9 (depurados) con línea de tendencia logarítmica



Si sabemos que el paso del tiempo afecta sensiblemente a las variables, a veces es preferible utilizar una línea de tendencia polinómica, buscando el orden o grado que más se adecúe a la distribución de puntos en el gráfico. Veamos un ejemplo.

En el gráfico 13 se representa la tasa de competencia oral (elevada al cuadrado) en una lengua menos extendida en una comunidad y en la tasa de uso de esa misma lengua. Si se trazan las líneas de tendencia polinómicas se puede observar fácilmente la armonía entre las trayectorias de ambas líneas (más suavizada en el caso de U). Por tanto se puede acreditar alguna relación entre las variables.

Gráfico 13. Tasas de competencia oral activa (al cuadrado,  $C^2$ ) y tasas de uso (U) de una lengua menos extendida en una comunidad





## 6. RELACIÓN DE MÁS DE DOS VARIABLES

### 6.1. Comparación de correlaciones

En una investigación sobre la matrícula en una universidad grande<sup>8</sup> se compara los porcentajes de estudiantes que abandonan (que no se matriculan en dos cursos seguidos) (digamos X) y los porcentajes de estudiantes que precisan más de cuatro años (la duración establecida de los grados) para concluir las carreras (digamos Y). Se establecen estos porcentajes en las distintas titulaciones o grados. Se calculan estos porcentajes según los tres tipos de acceso del estudiante a la universidad: en primer lugar, los estudiantes procedentes del bachillerato, que superan un examen denominado Prueba de Acceso a la Universidad (PA), los estudiantes que acceden desde la formación profesional, después de superar los Ciclos Formativos de Grado Superior (FP) y las diversas modalidades de exámenes especiales para mayores de 25 años, de 40 años y de 45 años (MA). A continuación se agrupan las titulaciones por áreas y se calcula, en cada caso, el valor de R. Los resultados se recogen en la tabla 10.

Tabla 10. Coeficientes de correlación entre abandono y prolongación de estudios en diversos grados, según área y modalidad de acceso del estudiantado.

Área	PA	FP	MA
Ciencias Humanas	0,540	-0,348	-0,779
Ciencias Sociales	0,429	-0,344	-0,618
Ciencias Naturales	-0,666	-0,531	-0,860
Ingenierías	-0,927	-0,262	-0,754

¿Qué podemos concluir de esta tabla? Obsérvese que en la columna PA, esto es, los estudiantes que acceden normalmente desde las Pruebas de Acceso, hay dos «estilos». En las Ciencias Humanas y en las Sociales hay correlaciones positivas: en los grados, si el porcentaje de abandonos aumenta, también lo hace el de los estudiantes que prolongan el tiempo de estancia, y viceversa. En las Ciencias Naturales y las Ingenierías sucede al revés (no se han considerado en el estudio las Ciencias de la Salud, que tienen duración distintas). Aquellas carreras donde hay más abandonos, mengua la proporción de estudiantes que precisan más tiempo para concluir los estudios y viceversa. A primera vista, estas son más, digamos,

---

<sup>8</sup> Universitat de València, cursos 2009/10-2014/15.

«selectivas» (si abandonan más, mengua la prolongación; si abandonan menos, aumenta la prolongación). Ahora bien, adviértase que ese carácter «selectivo» que podemos relacionar con correlaciones negativas, se extiende a todas las áreas en el caso de las otras modalidades de acceso.

## 6.2. Matrices de correlaciones

Cuando tenemos más de dos variables, podemos hacer una matriz de correlaciones por pares de variables. Imaginemos que tenemos cinco variables V, W, X, Y y Z. Entonces podemos establecer la siguiente matriz de correlaciones.

	V	W	X	Y	Z
V	$R_{VV}$	$R_{VW}$	$R_{VX}$	$R_{VY}$	$R_{VZ}$
W	$R_{WV}$	$R_{WW}$	$R_{WX}$	$R_{WY}$	$R_{WZ}$
X	$R_{XV}$	$R_{XW}$	$R_{XX}$	$R_{XY}$	$R_{XZ}$
Y	$R_{YV}$	$R_{YW}$	$R_{YX}$	$R_{YY}$	$R_{YZ}$
Z	$R_{ZV}$	$R_{ZW}$	$R_{ZX}$	$R_{ZY}$	$R_{ZZ}$

Ahora bien

-La correlación de una variable con ella misma siempre es 1, por lo que no es preciso calcular  $R_{VV}$ ,  $R_{WW}$ ,  $R_{XX}$ , etc.

-Como tampoco importa el orden de las variables, como ya se ha explicado,  $R_{wv} = R_{vw}$ ,  $R_{xv} = R_{vx}$ , etc.

Por ello, solo es necesario calcular los valores a un lado de la diagonal, por ejemplo:

	V	W	X	Y	Z
V		$R_{VW}$	$R_{VX}$	$R_{VY}$	$R_{VZ}$
W			$R_{WX}$	$R_{WY}$	$R_{WZ}$
X				$R_{XY}$	$R_{XZ}$
Y					$R_{YZ}$
Z					

En realidad, en una matriz de correlaciones de N variables solo hemos de calcular M correlaciones, según la fórmula sencilla:  $M = [N*(N-1)]/2$

A continuación veremos un ejemplo del uso de la matriz de correlaciones en la investigación educativa. Tomemos la siguiente tabla de valores relacionados con la educación y la sociedad.

Tabla 11. Variables relacionadas con la sociedad y la educación de Estados europeos<sup>9</sup>

	Abandono educativo temprano (%)	Puntuación Matemáticas (PISA 2012)	Puntuación Lectura (PISA 2012)	Puntuación Ciencias (PISA 2012)	Expectativa de escolarización (años)	Porcentaje Educación Superior(30-34 años) (%)	Aprendizaje permanente (%)	Inversión pública como porcentaje PIB (%)	Desigualdad social (ingresos quintil superior/quintil inferior)
Alemania	10,6	514	508	524	18,2	32,0	7,9	4,98	4,3
Austria	7,6	506	490	506	17,2	26,3	14,1	5,8	4,2
Bélgica	12,0	515	509	505	19,6	43,9	6,6	6,55	4,0
Chequia	5,5	499	493	508	18,1	25,6	10,8	4,51	3,5
Dinamarca	9,1	500	496	498	19,8	43,0	31,6	8,75	4,5
Eslovaquia	5,3	482	463	471	16,4	23,7	3,1	4,06	3,7
Eslovenia	4,4	501	481	514	18,5	39,2	13,8	5,68	3,4
España	24,9	484	488	496	17,9	40,1	10,7	4,82	6,5
Estonia	10,5	521	516	541	18,1	39,1	12,9	5,16	5,4
Finlandia	8,9	519	524	545	20,5	45,8	24,5	6,76	3,7
Francia	11,6	495	505	499	16,5	43,6	5,7	5,68	4,5
Holanda	8,8	523	511	522	19,1	42,2	16,5	5,93	3,6
Hungría	11,5	477	488	494	17,7	29,9	2,8	4,71	4,0
Irlanda	9,7	501	523	522	17,5	51,1	7,1	6,15	4,7
Italia	17,6	485	490	494	17,1	21,7	6,6	4,29	5,5
Letonia	10,6	491	489	502	17,9	37,2	6,9	4,93	6,5
Lituania	6,5	479	477	496	18,9	48,6	5,2	5,17	5,3
Luxemburgo	8,1	490	488	491	15,1	49,6	13,9	3,153	4,1
Polonia	5,7	518	518	526	18,3	39,1	4,5	4,94	4,9
Portugal	20,8	487	488	489	18,0	27,2	10,6	5,27	5,8
Reino Unido	13,6	494	499	514	16,6	47,1	15,8	5,88	5,0
Suecia	7,5	478	483	485	19,9	47,9	26,7	6,82	3,7

Con estos datos podemos establecer la siguiente matriz de correlaciones

<sup>9</sup> Fuente: Eurostat.

## Matriz de correlaciones

	Abandono	Matemáticas	Lectura	Ciencias	Expectativa	Superior	Aprendizaje	Inversión	Desigualdad
Abandono		-0,253	0,014	-0,168	-0,119	-0,146	-0,106	-0,084	0,682
Matemáticas			0,790	0,841	0,344	0,148	0,179	0,292	-0,248
Lectura				0,845	0,296	0,374	0,126	0,338	-0,028
Ciencias					0,353	0,280	0,171	0,242	-0,076
Expectativa						0,230	0,502	0,710	-0,175
Superior							0,361	0,401	-0,031
Aprendizaje								0,689	-0,260
Inversión									-0,181
Desigualdad									

Se han sombreado las celdas con valores superiores a 1/3. Dejando de lado las elevadas correlaciones de las materias PISA entre ellas, hay que destacar dos agrupamientos de correlaciones:

-Abandono educativo temprano y desigualdad.

-Inversión pública correlaciona con expectativa de escolarización y con aprendizaje permanente; en menor medida con enseñanza superior y con lectura.

Por otro lado, adviértase la ausencia de correlación entre abandono y resultados PISA.

La matriz de correlaciones nos permite profundizar en las lógicas subyacentes a un conjunto de variables.

### 6.3. Análisis factorial

También el análisis factorial permite establecer relaciones numéricas entre diversas variables. Para realizar un análisis factorial precisamos el programa especializado SPSS de IBM.

Lo que hace análisis factorial es construir una serie de variables ficticias (llamadas «componentes»). El componente 1 sería la variable que mejor reproduciría la variabilidad del conjunto de las variables reales. El componente 2 sería otra variable ficticia que pretendería explicar la variabilidad no explicada por el componente 1 y así hasta que se explique el 100% de la variabilidad. El análisis factorial establece además el coeficiente de correlación entre los componentes y las variables reales.

Haremos un cálculo con alguna de las variables de la tabla 11

Tabla 11. Ejemplo de análisis factorial

#### Abreviaturas de las variables

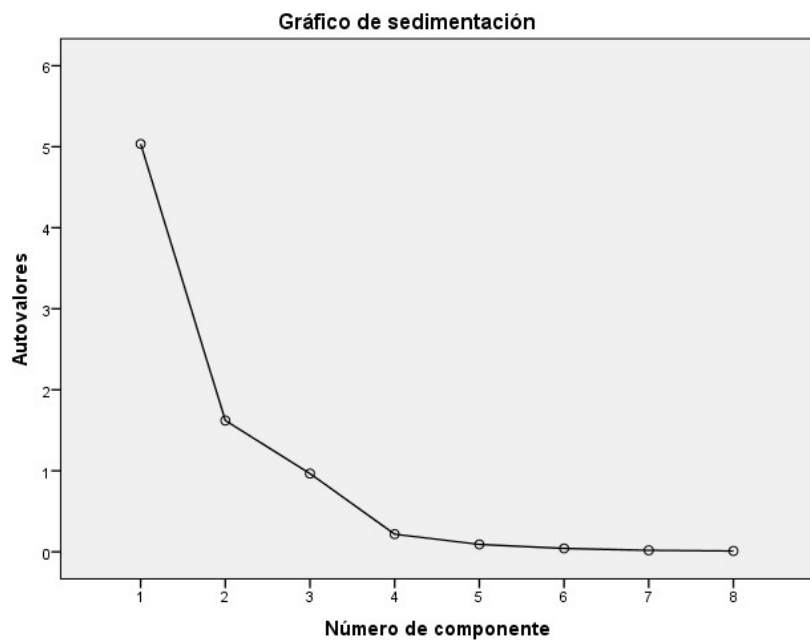
Matem_Total	Mean score in PISA 2012, Mathematics
Matem_Bajo	Share of low achievers (Below Level 2) in Mathematics PISA 2012
Lengua_Total	Mean score in PISA 2012, Reading
Ciencias_Total	Mean score in PISA 2012, Science
Exectativas	School Expectancy
Aprendizaje_H	Lifelong learning Male
Aprendizaje_D	Lifelong learning Female
Inversión	Annual expenditure per pupil

#### Varianza total explicada

Componente	Sumas de las saturaciones al cuadrado de la extracción	
	% de la varianza	% acumulado
1	62,930	62,930
2	20,255	83,185
3		

Matriz de componentes		
	Componente	
	1	2
Matem_Total	0,917	
Matem_Bajo	-,0941	
Lengua_Total	0,904	
Ciencias_Total	0,909	
Expectativas	0,651	
Aprendizaje_H	0,745	0,632
Aprendizaje_M	0,700	0,625
Inversión	0,442	0,663

Gráfico de sedimentación



El análisis factorial permite suponer un componente 1, que sería capaz de explicar el 62,9% de la variabilidad del grupo y que tendría una correlación de -0,941 con la variable Share of low Achievers (Below Level 2) in Mathematics PISA 2012 .

## II.- SOBRE LAS TÉCNICAS DE INVESTIGACIÓN CUALITATIVAS

---

### 1.- DEFINICIONES PRELIMINARES

Las técnicas de investigación cualitativas pretenden «interpretar», comprender el comportamiento de las personas. La interpretación se refiere a fragmentos discursivos (en expresión de Michel Foucault), tales como, la transcripción de una entrevista o de un grupo de discusión. Una interpretación no es una glosa, ni un pretexto. Una glosa o paráfrasis es volver a decir lo mismo con otras palabras, generalmente más. Una interpretación tampoco pretende emplear los fragmentos discursivos como pretexto para afirmar otra cosa distinta de lo que dicen. Interpretar no es considerar verdaderas las opiniones.

¿Como interpretar? Aconsejaba Theodor W. Adorno, uno de los precursores de las técnicas de investigación social cualitativas, que, cuando se interpreta, se debe renunciar al gesto del abogado que ha ganado un pleito. Interpretar no es ganar pleitos. Tampoco es jugar a detectives.

El antropólogo francés G. Condominas definía su investigación (el trabajo de campo etnográfico con un pueblo de las montañas del sudeste asiático) con una fórmula tan sencilla como acertada: hacer que lo exótico convierta cotidiano. Lo exótico, en definitiva, lo que no se entiende: ritos o tradiciones, el sentido de los que nos es opaco. Por el contrario, entendemos lo que nos es cotidiano, que pertenece a nuestro día a día. Para poder interpretar, hay que tener una cierta cotidianidad: entender la lengua, las costumbres, conocer la historia reciente de la comunidad, etc.

El historiador R. G. Collingwood acertó en explicar qué es interpretar en unos fragmentos de su *Autobiografía* que hablan sobre la ciencia de la historia, una disciplina donde la interpretación es fundamental: «La historia no consiste en saber qué acontecimientos siguieron otros. Significa introducirse en la cabeza de otras personas, mirar la situación con sus ojos, y pensar uno por su cuenta si la forma en la que habían abordado aquella situación era la debida ». Los fragmentos discursivos interpretan para introducirnos en la cabeza de otras personas, para ponernos en su piel, para tratar la situación con sus ojos. Por todo ello, afirmaba F. E. D. Schleiermacher, el padre de la ciencia de la interpretación (la llamada «hermenéutica»), que, en definitiva, «sobre mí mismo tengo que girar los ojos», para descubrir «la comunidad de los espíritus». Interpretar es descubrir la comunidad, hacerla cotidiana. Dotar de generalidad la indagación sobre lo subjetivo.

## **2.- ENTREVISTAS A INFORMADORES PRIVILEGIADOS**

Cuando hacemos una investigación sobre un asunto es recomendable entrevistar a personas que tengan conocimientos sobre el tema. Estas personas se suelen llamar «informadores privilegiados». Hay que precisar qué hay que hacer antes, durante y después de la entrevista, y explicar cómo interpretarla.

### **2.1.- Antes de la entrevista**

Supondremos que ya disponemos de un conocimiento del asunto para escoger la persona o personas que queremos entrevistar, es decir, que ya tenemos una «docta ignorancia», como diría el filósofo Nicolás de Cusa en la Edad Media. Lo primero que tenemos que hacer es anotar eventuales cuestiones y ordenarlas en un cuestionario provisional. Cuando se hace la entrevista, frecuentemente no se formulan todas las cuestiones o no en el mismo orden. Es lo que se llama una entrevista «semiestructurada». En todo caso, hay que estudiar el cuestionario y no dar la impresión de que solo lo leemos.

Antes de la entrevista hay que reflexionar sobre el lugar, la fecha y la hora adecuada de su realización. Es preferible hacer la entrevista en un lugar habitual de la persona entrevistada, en un espacio discreto y silencioso. En cuanto a la fecha y la hora, hay que prever también que la persona entrevistada tenga suficiente tiempo para dedicarnos. Antes de la entrevista, acordaremos el lugar, la fecha y la hora de su realización y, si es posible, haremos llegar un recordatorio a la persona entrevistada poco tiempo antes. También hay que comprobar el material que utilizaremos para registrar la entrevista y los accesorios (más adelante se volverá sobre ellos).

### **2.2.- Durante la entrevista**

En primer lugar, hay que pedir permiso a la persona entrevistada para registrar o grabar la entrevista. Incluso es conveniente disponer de un formulario (véase un modelo en el anexo 1). Si la persona no está de acuerdo en ser grabada, es mejor buscar otro informador privilegiado. No se deben grabar entrevistas con personas menores de edad y si lo hacemos (por ejemplo, estudiantado de los últimos cursos de Secundaria) debemos contar con la autorización de los padres o tutores. Siempre tenemos que hacer constar que no emplearemos datos o informaciones privadas. También es conveniente adquirir el compromiso de enviarle un borrador



del texto en el que aparecerá su aportación antes de que sea presentado. Por ello, hay que disponer de sus datos de contacto actualizados (teléfono y email).

Con el fin de grabar la entrevista es recomendable utilizar aparatos redundantes (es decir, más de un aparato), a fin de no perder el registro si falla una grabadora. Así, por ejemplo, se puede utilizar una cámara de vídeo (o una cámara fotográfica en modo grabación de imágenes) y un teléfono móvil. Antes debemos revisar los accesorios: baterías, pilas, memorias, trípodes, eventuales micrófonos, etc. Hay que tener en cuenta que si grabamos al aire libre, el viento entorpece mucho la comprensión del audio.

El hecho de registrar vídeo, y no sólo audio, presenta algunas ventajas. Frecuentemente, la gestualidad aporta informaciones complementarias importantes para el análisis de una entrevista. Por ejemplo, cuando una persona dice una cosa y hace con los dedos el gesto de ponerlo entre comillas; o cuando la gente con el rostro o las manos manifiesta escepticismo respecto de lo que ha dicho, por ejemplo, subiendo los hombros. Es preferible disponer la cámara a un lado, en un lugar discreto, enfocando la persona entrevistada en un plano medio. No se trata de poner la cámara «oculta», sino más bien lo que se pretende es que las declaraciones del informador no estén mediatizadas por el hecho de saberse grabado, que se olvide pronto de la presencia de los aparatos.

Es conveniente empezar la entrevista declarando el lugar, día, hora y nombre de la persona entrevistada y agradeciéndole de nuevo su participación. Esto sería el equivalente de la claqueta cinematográfica.

Debemos preparar algunas cuestiones para romper el hielo. Son preferibles preguntas generales, formuladas de una manera neutra, con las que el informador perciba que estamos interesados por lo que nos puede contar. Por ejemplo, si estamos haciendo una investigación sobre un aspecto de la educación, podemos preguntar sobre qué piensa de la enseñanza en general. Luego ya iremos formulando cuestiones que acoten el asunto que queremos averiguar. Cuando formulemos las cuestiones, huiremos de enunciados que orientan la respuesta en alguna dirección. Hay que preparar y ensayar las formulaciones de las preguntas de la manera más neutra posible.

Frecuentemente la persona entrevistada empieza a contestar con expresiones breves, habitualmente monosílabos. El entrevistador debe ser consciente de que está siendo analizado por el informador privilegiado, y hasta que no tenga una cierta confianza no aportará informaciones relevantes.

Es recomendable tomar notas, pero no hay que copiar a la letra lo que el entrevistado dice (porque puede dar una sensación demasiado inquisitiva y porque ya se analizarán después las respuestas con las grabaciones). Podemos escribir algunas anotaciones de manera no exhaustiva, con el fin, por ejemplo, de formular ulteriores cuestiones. En este sentido, hay que ser escrupuloso en utilizar exactamente las expresiones del informador por no orientar sus respuestas.

Es frecuente que las personas entrevistadas también plantean al entrevistador algunas cuestiones o pidan la opinión sobre lo que han declarado. El entrevistador deberá pensar con antelación como zafarse de este tipo de cuestiones. Un recurso, ante una pregunta directa, es decir que se contestará al finalizar la entrevista. Si hay que dar alguna respuesta durante la entrevista, es necesario que sea lo más neutra posible.

La entrevista no debe tener una duración excesiva. Si nos quedamos cortos, siempre podremos volver a entrevistar al informador. Si nos pasamos de tiempo y el informador se cansa, probablemente rechazará continuar colaborando.

Hay un momento muy importante en una entrevista: cuando le hemos agradecer la colaboración a la persona entrevistada y nos ve apagar los aparatos de grabación. En este momento, la persona entrevistada suele relajar el control que estaba ejerciendo sobre lo que declaraba, se desinhibe y puede hacer afirmaciones que había reprimido, y que iluminan bien su posición. Por ello es recomendable hacer tres cosas: 1º) Apagar ostensiblemente la cámara y comentarlo, pero mantener un cierto tiempo la segunda grabadora en funcionamiento (por ejemplo, el teléfono móvil), o al revés, de ese modo si se desinhibe y dice algo relevante lo habremos recogido con una grabadora; 2º) Si no lo podemos registrar, recordar bien lo que el entrevistado dice y hace cuando se produce la desconexión, y si es posible 3º) anotar todo, con la mayor fidelidad, cuando finalizamos la entrevista.

Es conveniente dar a la persona entrevistada un pequeño obsequio, en compensación por el tiempo que nos ha dedicado. Ejemplos de regalos pueden ser un pen-drive o un bonus de obsequio de algún establecimiento.

### **2.3.- Después de la entrevista**

Lo más pronto posible, cuando aún tenemos el recuerdo de la entrevista más fresco, procederemos a transcribir de manera completa o parcial (copiando aquellos fragmentos más significativos) (sobre cómo transcribir una entrevista véase más adelante). Copiaremos los ficheros con las entrevistas en un ordenador y tomaremos la precaución de hacer copias de seguridad y no borrar los de la cámara o el móvil hasta que no dispongamos de la transcripción. Así también servirán como copia de seguridad.

Además de la transcripción de lo grabado, hay que hacer un informe posterior a la entrevista, donde tenemos que hacer constar dónde y cómo se ha realizado, transcribimos las notas que hemos tomado y todo aquello que pueda orientar una interpretación posterior. De esta manera ya podemos proceder a la interpretación.

La interpretación de una entrevista a un informador privilegiado pretende extraer conocimientos sobre un determinado ámbito. Ahora bien, la obtención de estos conocimientos necesita de una serie de precauciones metodológicas, tales como:

- a) Mantener la fidelidad a lo que dijo el informador. No hay que sustituir la literalidad de lo que él ha expresado por una interpretación nuestra que cambie el sentido de sus palabras.
- b) Posteriormente hay que contrastar lo que dice mediante otras fuentes de información, porque no podemos a priori atribuirle verdad, de manera incondicional o ingenua a todo lo que dice un informador. En el periodismo se suelen utilizar, al menos, dos fuentes que ratifiquen un hecho.
- c) Hay que prestar atención además a la forma en que nuestro informador puede decirnos medias verdades o también evitar ciertos comentarios. A veces también resulta relevante lo no dicho. Las personas suelen comportarse de esta manera ante conflictos o situaciones incómodas. En definitiva, a todas las personas nos gusta presentarnos como coherentes.
- d) Después de hacer la interpretación, revisaremos la transcripción de la entrevista o las grabaciones, para comprobar que no hemos olvidado nada importante.

### **3.- ENTREVISTAS A PERSONAS EN GENERAL**

Frecuentemente se utilizan entrevistas a personas en general en la investigación social o educativa. En este caso no nos interesa tanto lo que la persona dice, sino más bien *cómo* lo dice. El presupuesto de esta metodología es que una persona cualquiera organiza sus explicaciones sobre los fenómenos sociales con discursos que capta del ambiente, y lo que la entrevista nos permite analizar es precisamente cómo son estos discursos.

En este caso, la persona está seleccionada no para que sea una informadora privilegiada, sino precisamente porque no mantiene una relación especial con el tema. Por ejemplo, podemos entrevistar a una serie de madres o padres de pequeños escolarizados para estudiar la participación en la escuela de las familias o sus preocupaciones; pero tendremos que prescindir, por ejemplo, de padres o madres que a su vez sean docentes.

Es importante advertir que no se trata de elegir un conjunto de personas que represente el conjunto en sus características (lo que se llama una muestra). De eso se trata en las técnicas cuantitativas, pero no en las técnicas cualitativas. Sabemos que hemos entrevistado suficientes personas no por *muestreo*, sino por *saturación*, es decir, para que comiencen a repetir lo que ya nos han dicho otros.

Como se prepara y realiza una entrevista a personas en general no presenta muchas diferencias de la entrevista a informadores privilegiados. La distinción fundamental se encuentra en la interpretación.

A continuación se añade un modelo de autorización.

*Autorización para la grabación en vídeo y grabación de voz*

*[Tratamiento] [NOMBRE Y APELLIDOS], con documento de identidad núm [Número], doy mi consentimiento a la [Institución o entidad que hace la investigación] para la grabación de secuencias vídeo o grabaciones de la voz que se realizan durante el evento de [Entrevista, grupo de discusión...] llevado a cabo en [Lugar] el [día] de [mes] de [año], dentro del proyecto de investigación [universitaria, académica...] sobre [asunto o tema].*

*Entiendo que el uso de la imagen o de la voz del participante, así como los datos obtenidos serán utilizados únicamente por el equipo de investigadores designados a tal fin y solo con fines estadísticos y de desarrollo de dicho proyecto de investigación, organizado por [Persona o grupo de personas que hace la investigación] de [Institución o entidad en la que se hace la investigación].*

*La autorización podrá ser revocada en cualquier momento, debiendo comunicarse a través de un correo electrónico a la dirección [email o dirección de la persona investigadora].*

*Esta autorización se aplica a las secuencias filmadas en vídeo o grabaciones de voz que se puedan recopilar como parte del desarrollo del evento y para los fines que se indican en este documento.*

*Tomando todo esto en consideración y en estas condiciones, CONSIENTO en participar en la grabación de las sesiones y que los datos que se derivan de mi participación sean utilizadas para cubrir los objetivos especificados en el documento.*

*En [lugar], [día / mes /año].*

*[Firma]*

*Firmado:*

#### **4.- GRUPOS DE DISCUSIÓN**

Los grupos de discusión o de debate son una técnica cualitativa habitual, que podríamos considerar una extensión de las entrevistas a personas en general. Se trata de formar un grupo de personas (lo ideal es entre 6 u 8 personas), que habitualmente no se conocen entre ellas y que tratan un asunto en un tiempo determinado (hora y media o dos horas, como mucho). El investigador actúa como moderador del debate, intentando que profundicen en aquellas líneas de argumentación en las que se enfrentan los discursos predominantes. También intenta que nadie monopolice el uso de la palabra o avasalle a los demás con su argumentación. Es conveniente en estos casos que actúen dos investigadores: uno que adopta un papel más activo (moderador) y otro que desempeña un papel más analítico, de observación de la dinámica del grupo.

En el caso de los grupos de discusión se precisa una sala adecuada (en algunos laboratorios de ciencias sociales se dispone de salas con registro de vídeo y audio discreto para grupos) y duplicar los aparatos de registro (es bueno que los aparatos que registran vídeo esten cruzados, para poder estudiar las reacciones de todos las personas participantes).

#### **5.- TRANSCRIPCIÓN E INTERPRETACIÓN**

##### **5.1. Transcripción**

En algunos contextos hay normas estrictas de transcripción, que hay que seguir. Si no se dispone de estas normas, se puede proceder del modo siguiente.

Se propone transcribir una entrevista utilizando una tabla con tres columnas y muchas filas. En la fila de la izquierda haremos algunas indicaciones del tiempo de la grabación. De esta manera podremos localizar rápidamente un pasaje de la transcripción en el registro audiovisual o viceversa. En la columna central (mucho más ancha que las otras dos) haremos la transcripción del texto. Cada vez que haya una pregunta, una respuesta o un cambio en la respuesta (un punto y aparte), utilizaremos otra fila. No es necesario que en cada fila anotamos el tiempo de la grabación. Es conveniente hacerlo, al menos, una vez por página. En la columna de la derecha anotaremos observaciones al texto de la segunda columna. Por ejemplo:

núcleos significativos, expresiones interesantes, etc. Esto nos ayudará después a organizar la interpretación.

En las transcripciones hay que anotar no sólo las frases que emiten las personas entrevistadas, sino también todo aquello que ayude a su comprensión: énfasis, tono irónico, gestos, silencios, etc. En el caso de los silencios, podemos hacer constar, incluso, su duración. Todo esto se puede anotar en la transcripción utilizando paréntesis o llaves y letra cursiva, por ejemplo: *{en tono irónico}*, *{con mucho énfasis}*, *{silencio 3 segundos}*, etc. También debemos anotar una transcripción dudosa: *{palabra dudosa}*, *{dudoso, tal vez dice...}*.

No se deben sustituir las palabras, pero podemos corregir aquellas que son propias de la lengua oral, pero no de la lengua escrita. Por ejemplo, si una persona dice: «*Es una maestra muy prepará*», podemos transcribirlo como «*Es una maestra muy preparada*», cuando suponemos que la persona entrevistada lo hubiera hecho así si hubiera escrito su frase.

En las transcripciones de entrevistas a personas en general o grupos de discusión utilizaremos nombres ficticios. Habitualmente se emplean nombres propios, siguiendo el orden alfabético, tanto en hombres como en mujeres. Por ejemplo: Aurora, Berta, Clara, Daniela, Elsa, etc. En el informe final, haremos constar alguna característica. Por ejemplo: Aurora (maestra de Educación Infantil), etc. Pero no hay que dar datos que impliquen la identificación unívoca de la persona.

## 5.2. Interpretación

En el caso de entrevistas a informadores privilegiados, se trata de obtener conocimientos contrastados.

En el caso de una entrevista a una persona en general, la interpretación no debe consistir en hacer un repertorio de sus opiniones, sino que debe centrarse en mostrar la forma en que utiliza los significados. Supondremos que la persona organiza lo que dice a partir de «discursos» que toma de su entorno, que es como un altavoz que recoge varias versiones de la interpretación de los hechos que estamos estudiando. Pero las explicaciones que da no son monolíticas, sino que se componen con fragmentos de unos discursos u otros. La función de la interpretación es distinguir estos fragmentos discursivos. Para ello, hemos de tener en cuenta:

a) La articulación de diversos significados para identificar los fragmentos discursivos. Resulta recomendable hacer un mapa conceptual, donde haremos constar las palabras que presentan los núcleos de significado más importantes y

las relaciones mutuas. Hay que hacerlo con fidelidad a las expresiones de la persona entrevistada.

Un ejemplo. En un grupo de discusión, unas madres hablan de su relación con los maestros utilizando reiteradamente el concepto de «escudo».

Yo he intentado hablar para esto con el equipo directivo y nos dicen... lo primero el escudo {posa una mà oberta al davant, com un escut}. (Aurora, madre) [...]

Y todo el mundo se escuda en la legislación. (Aurora) [...]

Lo que creo es que [...]a lo mejor, profesores y padres se podrían mostrar más. «Mira, yo voy a trabajarlo de esta manera». Ni que un profesor se tenga que escudar en un libro, para que nadie le diga nada. (Berta, madre)

Cuando esas personas repiten esa noción y la presentan reiteradamente ante el resto del grupo, que no discrepa, vamos más allá de opiniones personales y nos encontramos ante representaciones generales sobre las relaciones entre docentes y familias.

b) La presencia de paradojas o contradicciones, es decir, aquellos pasajes en los que la persona entrevistada dice una cosa y su contraria, prácticamente sin darse cuenta. En los aforismos del filósofo griego Heráclito tenemos muchos ejemplos, tales como: «El camino hacia arriba y el camino hacia abajo es uno y el mismo». Es lo que llamamos «sinalaxis». Una muy frecuente es cuando una persona dice: «yo no soy racista, pero...». Las sinalaxis nos indican los lugares donde los fragmentos discursivos se enfrentan.

Veamos un par de ejemplos relativos a entrevistas a estudiantado de último curso de bachillerato en relación a la prueba de acceso a la universidad y las carreras en las que piensan matricularse. En el primer caso, la persona declara al mismo tiempo que tomará la decisión «después» aunque ya la tomado antes porque elige lo «que me puede llegar»:

Pregunta: Y de qué depende... Vamos a ver, la pregunta es la siguiente, vamos a ver, ¿piensas decidirte por una ingeniería antes del selectivo, digamos?

Respuesta: Después.

P: O va a depender...

R: De la nota que saque, pues...

P: Según la nota que saques te vas a orientar a una u otra.

R: Sí, en principio, a ver, la aeroespacial yo creo que no me va a llegar. Entonces la industrial es la que...

P: ¿Hay aeroespacial en el politécnico? Es que no lo sé.

R: Creo que sí. Entonces la industrial es la que sí pienso que me puede llegar, entonces...

(Clara, estudiante de 2º de bachillerato, Centro público)

En esta misma entrevista se puede apreciar como una contestación inicial afirmativa a propósito de si sabe qué va a hacer se convierte en negativa poco después.

Pregunta: ¿Y tú crees que podrás hacer lo que quieres?

Respuesta: Sí, ¿no? Yo creo que sí.

P: ¿Qué nota dices?

R: No sé cual, a lo mejor ocho, diez, once... No lo sé.

P: Bueno, tendrás una buena media tú, ¿no?

R: De primero nueve con cuatro, pero de segundo a ver si puedo sacar más o menos similar, y en el selectivo pues que se pueda sacar simplemente, y con eso yo creo que ya...

P: Ya puedas hacerla.

R: Espero.

P: Y si finalmente las cosas se complican, ojalá no, ¿tienes un plan B?

R: No, o sea... ¿Que se complicaran que no pudiera llegar a la nota? No sé, me metería a lo mejor en otra carrera así... Pensaba en una carrera de idiomas o así, pero es que también tiene la nota altita, entonces no sé.

(Clara, 2º Bachillerato, Centro público)

Lo pertinente aquí es que se desvela el contexto de incertidumbre y cómo los individuos perciben el carácter de profecías autocumplidas de los sistemas de selección académica.

c) Los usos lingüísticos concretos. Hay que analizar, por ejemplo, el uso de diminutivos, palabras en otra lengua, frases hechas, reiteraciones, tautologías («un profesor es un profesor»). En una investigación de Marina Subirats, catedrática en la Universitat Autònoma de Barcelona, se constató cómo los y las docentes se referían a los trabajos escolares de las estudiantes empleando diminutivos con mayor frecuencia que cuando trataban de los trabajos de los estudiantes.

Un ejemplo. En un grupo de discusión sobre multilingüismo en la escuela, una madre comenta a propósito de la normativa educativa:

[...] la persona o el ente *{rie la persona que está al lado}* que se pone a desarrollar una normativa... El ente *{rien todas las personas}*... en el papel puede ser muy bonita, pero no pide realmente consenso, opinión, experiencia, de los que realmente están con los niños.  
(Daniela, madre)

Lo pertinente para la interpretación en este fragmento es, por un lado, la reiteración de dos términos: «ente» y «realmente». El primer término es inhabitual y denota lejanía, extrañamiento: la administración educativa se presenta como algo ajeno, remoto. Y esta caracterización es recibida por el resto del grupo con risas de asentimiento. Por otro lado, la contraposición con lo que «realmente» se demanda o sucede. Se reafirma así una contraposición (el ente/lo real) con la aprobación general, que va más allá de la opinión de Aurora porque indica la representación predominante.

d) Como siempre, hay que prestar atención a lo no dicho, a aquellas cosas que la persona entrevistada evita en su declaración. En este punto hay mucho riesgo de sobreinterpretación.



La interpretación de los grupos de discusión suele seguir las mismas pautas que las entrevistas a las personas en general, aunque en ellos resulta más fácil la identificación de los diversos fragmentos discursivos en conflicto.

Hay que recordar nuevamente que en la publicación definitiva de la investigación es preciso cambiar el nombre de las personas para mantener el anonimato, así como evitar cualquier circunstancia que pueda servir para identificar a ella o a otras personas no publicas que alude. Si por ejemplo una persona habla de otra sustituiremos los nombres propios por un paréntesis cuadrado con una indicación. Por ejemplo: [Nombre de otro docente], [Nombre de un centro educativo de la localidad], [Nombre de un alumno], etc.

## **6.- INVESTIGACIÓN BIOGRÁFICA**

La investigación biográfica es una técnica cualitativa muy útil en la indagación de asuntos educativos. Sobre este asunto, se remite al texto *Fundamentos filosóficos y sociológicos de la investigación biográfica*, disponible en internet en la dirección.

<http://www.uv.es/fjhernan/uab.pdf>