Universidad Zaragoza

1542

Trabajo Fin de Máster

Cálculo e interpretación del coeficiente de correlación lineal

Computation and interpretation of the linear correlation coefficient

Autora

Alicia Gracia Cano

Director

Antonio González Herrera

Facultad de Educación

Curso 2017/2018

ÍNDICE

1. EL OBJETO MATEMÁTICO A ENSEÑAR

1.1 Objeto matemático a enseñar

El objeto matemático a enseñar es el coeficiente de correlación, cuyos aspectos a tratar son: definición y propiedades, significado y aspectos interpretativos junto con sus aplicaciones.

1.2 Situación del currículo

El estudio de la relación entre dos variables y la posible predicción de una de ellas a partir de otra, es decir, de la correlación y regresión, es fundamental en estadística y en investigación empírica. En primer lugar, por su relación con muchos conceptos estadísticos fundamentales, como son los de variación, distribución, centralización o de dispersión; en segundo lugar, porque su estudio se vincula también a la toma de decisiones, y por último, porque permite superar la tendencia a emplear estrategias intuitivas incorrectas de estimación o asociación con causalidad (Estepa, 2008; Zieffler y Garfield, 2009).

La Orden ECD/494/2016, de 26 de mayo, por la que se aprueba el currículo del Bachillerato y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón, configura el cálculo e interpretación de la correlación en el primer curso de Bachillerato dentro del bloque 5 "Estadística y Probabiliad" para las matemáticas I y dentro del bloque 4 para las matemáticas aplicadas a las ciencias sociales:

Estadística descriptiva bidimensional.

Tablas de contingencia.

Distribución conjunta y distribuciones marginales. Medias y desviaciones típicas marginales. Distribuciones condicionadas.

Independencia de variables estadísticas.

Estudio de la dependencia de dos variables estadísticas. Representación gráfica: nube de puntos.

Dependencia lineal de dos variables estadísticas. Covarianza y correlación: Cálculo e interpretación del coeficiente de correlación lineal.

Regresión lineal. Estimación. Predicciones estadísticas y fiabilidad de las mismas (pp 14107-14108).

Cabe señalar que también figura como "introducción a la correlación" en el currículo de matemáticas orientadas a las enseñanzas académicas del 4º curso de ESO (contemplado en la Orden ECD/489/2016, de 26 de mayo, por la que se aprueba el currículo de la Educación Secundaria Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón) dentro del bloque 5 "Estadística y Probabilidad":

Introducción a la combinatoria: combinaciones, variaciones y permutaciones.

Cálculo de probabilidades mediante la regla de Laplace y otras técnicas de recuento.

Probabilidad simple y compuesta. Sucesos dependientes e independientes.

Experiencias aleatorias compuestas. Utilización de tablas de contingencia y diagramas de árbol para la asignación de probabilidades.

Probabilidad condicionada.

Utilización del vocabulario adecuado para describir y cuantificar situaciones relacionadas con el azar y la estadística.

Identificación de las fases y tareas de un estudio estadístico.

Gráficas estadísticas: Distintos tipos de gráficas. Análisis crítico de tablas y gráficas estadísticas en los medios de comunicación. Detección de falacias. Medidas de centralización y dispersión: interpretación, análisis y utilización.

Comparación de distribuciones mediante el uso conjunto de medidas de posición y dispersión.

Construcción e interpretación de diagramas de dispersión. Introducción a la correlación (p. 13075).

En este TFM abordamos el coeficiente de correlación contemplado en 1º de Bachillerato (Matemáticas I).

1.3 Campos de problemas, técnicas y tecnologías

La propuesta didáctica del cálculo e interpretación de la correlación va a ser expuesta mediante una serie de campos de problemas planteados de forma gradual conforme los cuatro niveles superiores en la taxonomía de Bloom (Bloom, 1956), desde la comprensión del objeto matemático hasta su evaluación.

En un primer momento, se recuerdan los conceptos ligados a dependencia entre variables que los alumnos tienen del curso anterior en la sesión inicial. Posteriormente, mediante el campo de problemas 1 se crea la necesidad de “definición del coeficiente de correlación” para informar de la intensidad y sentido de la dependencia lineal entre variables. A continuación, en el campo de problemas 2, se conecta con el sentido gráfico de la correlación permitiendo desarrollar una comprensión más significativa de la correlación. Por último, en los campos de problemas 3 y 4 es cuando se expone situaciones problema para incidir en la toma de decisiones y aspectos interpretativos respectivamente.

Las técnicas y tecnologías asociadas a cada campo de problemas son:

Campo de problemas	**Técnicas**	**Tecnologías**
CP1: Aproximación a la definición de correlación	TC1.1: Dibujo de nube de puntos a partir de tabla de doble entrada. TC1.2: Cálculo de posibles valores de la covarianza y correlación a partir de una nube de puntos. TC1.3: Predicción de la correlación según una nube de puntos.	TG1.1: Relación entre la nube de puntos con la intensidad y sentido de la correlación. TG1.2: Coeficiente de correlación como la versión adimensional de la covarianza. TG1.3: Propiedades de la covarianza. TG1.4: Relación de la distancia entre el centro de gravedad de una nube de puntos y los datos atípicos con el valor de la correlación.

CP2: Propiedades de la correlación.	TC2.1: Deducción de la correlación a partir de coordenadas cartesianas. TC2.2: Deducción de la correlación según la distancia de los puntos al centro de gravedad.	TG2.1: Relación entre la nube de puntos y mínimos cuadrados. TG2.2: Definición e interpretación gráfica de la covarianza. TG2.3: Relación entre el coeficiente de correlación y el centro de gravedad de una nube de puntos.
CP3: Correlación para tomar decisiones.	TC3.1: Uso de la calculadora para hallar la correlación a partir de una tabla de doble entrada.	TG3.1: Definición de la correlación como valor cuantitativo de la relación entre dos variables. TG3.2: Propiedades del coeficiente de correlación.
CP4: Interpretación del coeficiente de correlación.	TC4.1: Cálculo del coeficiente de correlación. TC4.2: Interpretación del coeficiente de correlación.	TG4.1: Propiedades de la correlación. TG4.2: Definición operacional del coeficiente de correlación

2. ESTADO DE LA ENSEÑANZA-APRENDIZAJE DE LA CORRELACIÓN

2.1 Introducción escolar del objeto matemático

Para estudiar la introducción escolar del objeto matemático "cálculo e interpretación de la correlación", he utilizado el libro de Matemáticas I del curso 1º de Bachillerato editado por Editex; en concreto, la unidad didáctica 16 "*Distribuciones* *bidimensionales. Correlación y regresión*". Aunque en el centro donde he realizado mi Practicum no cuenta con los cursos de 1º y 2º de Bachillerato, es un texto de uso frecuente en la mayoría de los centros.

Al inicio del tema, propone unas cuestiones iniciales como primer contacto de lo estudiado en el curso anterior tales como la media, desviación típica, nube de puntos e idea intuitiva de una posible correlación a través de la representación gráfica de un conjunto de datos. Estas cuestiones iniciales se proponen a través de una serie de tres ejercicios en los que ya se muestra una tabla resumen de datos.

Junto con estas cuestiones iniciales, el texto completa esta introducción con un "mapa conceptual" sobre el estudio de la estadística:

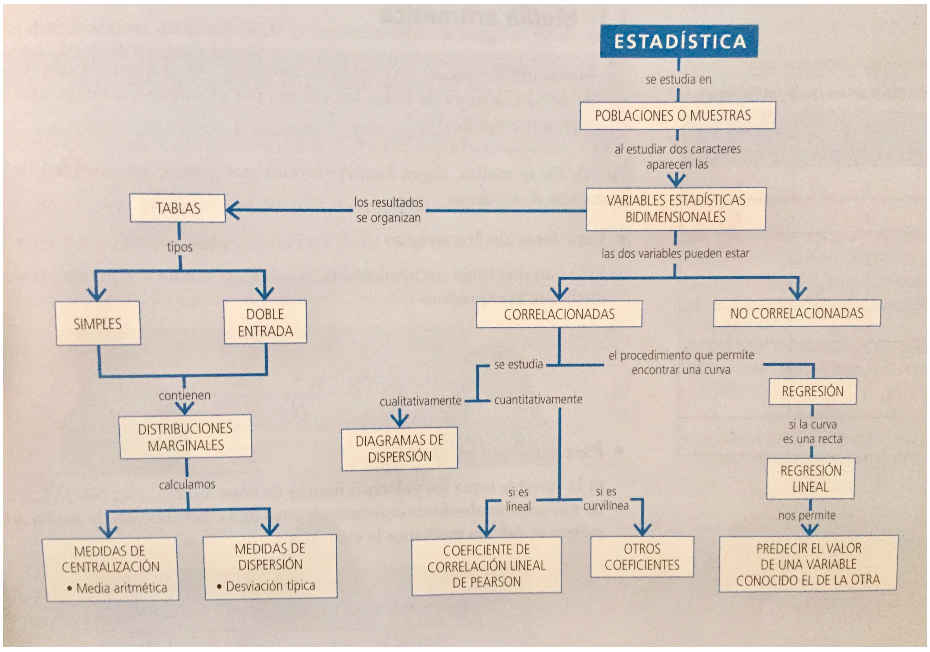


Figura 1. Mapa conceptual de la unidad 16: distribuciones bidimensionales. Correlación y regresión. Libro Matemáticas I. 1º bachillerato. Ed. Editex.

Conviene hacer notar la inexactitud, o al menos la falta de completitud, en esta presentación de los conceptos expuestos sobre correlación en las variables estadísticas bidimensionales. Si bien el concepto de dependencia estadística o aleatoria en ocasiones se identifica con correlación, la correlación se limita al tratamiento de variables cuantitativas (Batanero, Gea, López-Martín, Arteaga, 2017). En el mapa conceptual no se representa tal distinción.

En concreto, en el análisis del texto que nos ocupa, el concepto de correlación se presenta de manera formal a continuación de un breve acercamiento sobre "dependencia o correlación” tal y como se titula el apartado anterior. Como indica Sánchez Cobo (1999), el estudio de conceptos ligados a la dependencia estadística es un paso previo al análisis de la regresión, ya que la existencia de dependencia entre variables es la motivación en la búsqueda de modelos que permitan expresar relaciones funcionales entre las mismas.

En nuestro estudio, los conceptos que se proponen en el texto analizado son siempre referidos a la disposición de la nube de puntos:

- Dependencia funcional, si la nube de puntos se sitúa en la gráfica de una función, excepto que sea constante.

- Dependencia lineal, si la nube de puntos se sitúa sobre una recta.

- Correlación o dependencia aleatoria, si la nube de puntos se sitúa próxima a la gráfica de una función. El grado de correlación puede ser fuerte, débil, positivo o negativo.

- Independencia o ausencia de correlación (en este caso, simplemente lo acompaña de una representación gráfica).

Cabe volver a destacar la confusión que puede generar al emplear como sinónimos los conceptos de dependencia aleatoria y correlación.

En el siguiente apartado, “correlación lineal”, se presenta la expresión para el calculo del coeficiente de correlación lineal de Pearson. Junto con la escala de valores del coeficiente se exponen una serie de ejemplos sobre dependencia funcional, dependencia estadística o correlación positiva (vuelve a emplearlo como sinónimos), correlación negativa e independencia estadística.

Hacemos notar que la noción de “independencia” no es definida en ningún momento, quizá motivado por su fácil visualización en el diagrama de dispersión. Aún así, señalamos la pertinencia de definir este concepto adecuadamente puesto que es la base de muchos temas estadísticos posteriores; por ejemplo, en inferencia, un supuesto básico de aplicación de la mayor parte de contrastes estadísticos es admitir la independencia estadística de los datos de la muestra.

Respecto a la forma de presentar la fórmula del cálculo del coeficiente de correlación lineal también señalaremos que, aunque si bien el texto acompaña esta expresión junto a un ejemplo de tabla de doble entrada, sólo utiliza un sumatorio sencillo.

Por último, cabe observar que en este texto no se utiliza el recurso didáctico propuesto en Holmes (2001) consistente en la división en cuatro cuadrantes de la nube de puntos, utilizando rectas paralelas a los ejes que pasan por el centro de gravedad formado por las medias de cada variable. Puesto que a lo largo de las definiciones se alude al recurso de la nube de puntos, mostrar esta representación puede ayudar a los alumnos a desarrollar una comprensión más significativa de la covarianza y su relación con el coeficiente de correlación de Pearson.

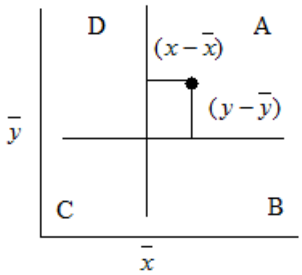


Figura 2. Distancia a la media de cada dato (Holmes, 2001, p.68).

De este modo, y realizando los productos que corresponden a cada valor se obtiene una tabla como la que se muestra en la Figura 3:

	A	B	C	D
$(x-\bar{x})$	$+$	$+$	$-$	$-$
$(y-\bar{y})$	$+$	$-$	$-$	$+$
$(x-\bar{x})(y-\bar{y})$	$+$	$-$	$+$	$-$

Figura 3. Estudio del signo del producto de la diferencia de medias.

Llamando “sd” a la desviación típica de cada variable y dividiendo todos esos productos entre el total de datos, implica el siguiente cálculo que facilita el coeficiente de correlación de Pearson:

$$\frac{1}{n}\frac{\sum(x-\bar{x})\cdot(y-\bar{y})}{sd(x)sd(y)}=\frac{\sum(x-\bar{x})\cdot(y-\bar{y})}{\sqrt{\sum(y-\bar{y})^2\sum(x-\bar{x})^2}}$$

2.2 Campos de problemas, técnicas y tecnologías que se enseñan habitualmente

Los problemas que se trabajan en el cálculo de la correlación se fundamentan en la mera obtención del coeficiente correlación de Pearson, el cual es definido en los textos generalmente desde el punto de vista operacional y acompañado de ejemplos que, en su mayoría, implícitamente, ofrecen una definición estructural del mismo (Batanero, Gea, López-Martín y Arteaga, 2017). Así, en el libro en el cual nos apoyamos se indica:

> Con la observación de la nube de puntos correspondiente a una distribución (X,Y) podemos identificar la correlación existente entre las variables X e Y. Además, podremos cuantificar esta correlación utilizando coeficientes; en particular, para cuantificar la correlación de tipo lineal usaremos el coeficiente de correlación lineal de Pearson, cuya expresión es (p.384):

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

Las técnicas que se enseñan son: cálculo del coeficiente de correlación lineal de Pearson conforme la expresión anterior, junto con las expresiones del cálculo de la covarianza o varianza conjunta, σ_{xy}, y de las desviaciones típicas de las variables marginales X e Y, $\sigma_x \sigma_y$.

En el libro estudiado, estas técnicas no están convenientemente justificadas, simplemente se presenta su definición y se realizan actividades resueltas en donde se verifica su funcionalidad en base a la validez de los resultados. Además, cabe señalar la inexactitud de la notación empleada, puesto que se limita a usar sólo un sumatorio en el caso de la covarianza, y la posible confusión que puede generar en los alumnos el uso de "σ" para referirnos a la desviación típica en lugar de "S".

En cuanto a la deducción de estas fórmulas de cálculo, no se desarrollan ni se presenta el método sustentado en un desarrollo gráfico.

Las situaciones problemáticas que se proponen en el texto presentan un mayor número de actividades donde se utiliza la relación lineal directa en comparación con la inversa.

2.3 Efectos producidos por esta enseñanza en el aprendizaje del alumno

Emitir juicios de asociación efectivos en la toma de decisiones implica, en particular, el dominio de las nociones de correlación y asociación. Estas nociones no son tratadas como contenidos de enseñanza específicos en la educación obligatoria, aunque, implícitamente, se encuentran presentes en el estudio de las tablas de contingencia, que forma parte del contenido de la asignatura de Matemáticas en cuarto de ESO en sus dos opciones: A y B (Gea, Batanero, Arteaga, Cañadas, Contreras, 2014). En concreto, en la Comunidad Autónoma de Aragón se hace presente la relevancia y el sentido educativo de estas nociones, implícitamente, a través de los siguientes estándares de aprendizaje evaluables:

- Distingue la dependencia funcional de la dependencia estadística y estima si dos variables son o no estadísticamente dependientes mediante la representación de la nube de puntos.

- Cuantifica el grado y sentido de la dependencia lineal entre dos variables, mediante el cálculo e interpretación del coeficiente de correlación lineal.

- Calcula las rectas de regresión de dos variables y obtiene predicciones a partir de ellas.

- Evalúa la fiabilidad de las predicciones obtenidas a partir de la recta de regresión mediante el coeficiente de determinación lineal.

Esta secuencia de estándares de aprendizaje evaluables junto con la definición explícita de los conceptos en los libros de texto, nos lleva a observar un predominio de la presentación de la correlación puramente operacional sin prestar mucha atención al significado y aspectos interpretativos.

Autores como Ouvrier-Buffet (2004) han mostrado las dificultades que tienen los estudiantes para realizar definiciones o comprender definiciones de los conceptos y, por lo tanto, es importante que los libros de texto presentes definiciones asequibles y completas. A estas observaciones hemos de añadir (Estepa y Batanero, 1995):

> - La concepción determinista de la asociación. Los alumnos tienden a asignar un único valor de la variable independiente a cada uno de los valores considerados de la variable dependiente. Esto es, la relación de las variables sólo es considerada desde un punto de vista funcional.

- La concepción local de la asociación. Los alumnos utilizan parte de los datos del estudio y se limitan a confirmar la asociación según un subconjunto de éstos que de algún modo justifique algún tipo de patrón, obviando la tendencia global de los datos (p.163).

Atendiendo igualmente a la intensidad y dirección de la correlación, a la posible linealidad, y a las teorías previas de los alumnos, así como a la precisión de las estimaciones, las estrategias empleadas y la identificación de situaciones reales en que se presente un valor dado del coeficiente de correlación, a continuación se resumen los principales resultados que obtienen según Sánchez Cobo, Estepa y Batanero (2000):

- Los alumnos han mostrado una buena capacidad de estimación de la correlación, aunque en general, las tareas no siempre resultan sencillas, haciéndose más viable cuando mayor es la intensidad de ésta.

- Los alumnos comprenden con facilidad la adimensionalidad del coeficiente de correlación, corroborando los resultados de una investigación anterior por Sánchez Cobo (1999), así como la relación entre el signo de la correlación y el sentido en que covarían los valores de las componentes de una variable bidimensional; no obstante, presentan dificultades en la concepción de una covariación negativa, captada por menos del cincuenta por ciento de los sujetos.

- La estimación es más precisa al estimar el coeficiente de correlación a partir de un diagrama de dispersión así como la tarea inversa, haciéndose mayor esta precisión cuando la correlación es más intensa, corroborando con ello las investigaciones desarrolladas con anterioridad (Sánchez Cobo, 1999), donde los errores son mayores al construir una nube de puntos a partir de una descripción verbal y estimar el coeficiente de correlación desde una tabla de valores numéricos. (p. 297)

Todos estos resultados ponen de manifiesto la complejidad de estas nociones, incluso al finalizar la secuencia de enseñanza. Todo esto unido a la escasez de definiciones que de tipo instrumental-relacional incluyen los textos (siendo en su mayoría de tipo instrumental) provoca que se pueda "*transmitir una visión de las matemáticas como disciplina conformada por una colección de reglas y hechos que deben ser recordados y que se refieren sobre todo al cálculo*" (Sánchez Cobo, 1999, p. 287). Este mismo hecho lleva, según el mismo autor, a que *"algunos alumnos confunden los coeficientes de correlación y de determinación"* (Sánchez Cobo, 1999, p. 211). Además, las demostraciones se presentan con una marcada función explicativa y de convicción, obviando la necesidad de desarrollar en nuestros alumnos la capacidad de argumentar de un modo lógico (Gea, Batanero, Cañadas y Contreras, 2013).

Sería también importante desarrollar la comprensión de la diferencia entre correlación y causalidad mediante la discusión con los estudiantes sobre la explicación de la correlación, que no siempre es debida a una relación causa-efecto (Gea, Batanero, Roa, 2014). Otras posibles explicaciones sugeridas por Barbancho (1973) y citadas por Batanero, Gea, López-Martín y Arteaga (2017) son las siguientes:

> - Las variables pueden ser interdependientes, como en el caso de la longitud de piernas y la altura de una persona.
>
> - La existencia de una tercera variable que determine la correlación, esto es, las variables muestran dependencia, pero es indirecta.
>
> - Covariación casual o espúrea, cuando parece que en la covariación de dos variables hay cierta sincronía, lo cual podría interpretarse como la existencia entre ambas; sin embargo, ésta es causal o accidental. (p.64)

3. LOS CONOCIMIENTOS PREVIOS

3.1 Conocimientos previos necesarios

En cuanto al contenido de la noción de correlación, la investigación de Lavalle, Micheli y Rubio (2006) precisa los conceptos que necesitan ser incorporados en la enseñanza y aprendizaje de este objeto para favorecer su comprensión, así como los procedimientos con los que se vincula. Estas nociones y procedimientos que los alumnos deben haber adquirido a lo largo de los cursos anteriores son:

- *Variable estadística*, tanto unidimensional como bidimensional: variable explicativa (X) y variable explicada/respuesta (Y); los procedimientos son:

1) Identificar las variables en estudio.

2) Distinguir entre una variable respuesta y una variable explicativa.

3) Identificar las unidades de medida.

4) Calcular la media, la desviación típica, la varianza y la covarianza.

- *Datos bidimensionales*, que se tratan para ser comparados, analizados e interpretados; los procedimientos son:

1) Identificar los valores de las variables como pares ordenados.

2) Graficar la nube de puntos/diagrama de dispersión.

3) Identificar si hay relación entre las variables a partir del gráfico (existe o no relación, es lineal o no, etc.) e indicar aproximadamente qué tipo y grado de relación lineal existe (directa, inversa).

Asimismo, para facilitar el aprendizaje de los estudiantes sería necesario asegurarnos de una correcta presentación de los conceptos y sus propiedades, puesto que, tal y como señala el estudio de Estepa (2007), muchos alumnos confunden conceptos o tienen dificultades en la realización de tareas llevadas a cabo con los mismos.

3.2 La enseñanza anterior

La enseñanza anterior ha dado lugar a la adquisición de estos conocimientos, en 4º curso de Educación Secundaria Obligatoria, en la asignatura Matemáticas orientadas a las enseñanzas académicas, según la Orden ECD/489/2016, de 26 de mayo, por la que se aprueba el currículo de ESO y se autoriza su aplicación en los centros docentes de la Comunidad de Aragón.

En el bloque 5, llamado Estadística y Probabilidad, se presentan los siguientes contenidos:

- Introducción a la combinatoria: combinaciones, variaciones y permutaciones.

- Cálculo de probabilidades mediante la regla de Laplace y otras técnicas de recuento.

- Probabilidad simple y compuesta. Sucesos dependientes e independientes.

- Experiencias aleatorias compuestas. Utilización de tablas de contingencia y diagramas de árbol para la asignación de probabilidades.

- Probabilidad condicionada.

- Utilización del vocabulario adecuado para describir y cuantificar situaciones relacionadas con el azar y la estadística.

- Identificación de las fases y tareas de un estudio estadístico.

- Gráficas estadísticas: Distintos tipos de gráficas. Análisis crítico de tablas y gráficas estadísticas en los medios de comunicación. Detección de falacias. Medidas de centralización y dispersión: interpretación, análisis y utilización.

- Comparación de distribuciones mediante el uso conjunto de medidas de posición y dispersión.

- Construcción e interpretación de diagramas de dispersión. Introducción a la correlación.

3.3 Verificación de conocimientos previos

Previo a cualquier experiencia de aula, Reid y Pecotcz (2002) aconsejan realizar un estudio previo que permita informar de las concepciones que los estudiantes tienen acerca de la Estadística.

En la primera sesión se planteará una práctica de aproximación, que nos permitirá adaptar nuestra programación de enseñanza y aprendizaje al entorno de los alumnos con objeto de ampliar sus conocimientos y comprensión en cuanto a la Estadística y, en concreto, a la correlación.

El objetivo de esta primera sesión es acercar a los alumnos la terminología estadística, así como facilitar la comprensión del lenguaje y conceptos estadísticos referidos a la correlación. Para ello, presentaremos una actividad en el aula en la que los alumnos podrán comprobar la diferencia entre relación funcional y relación estadística.

SESIÓN INICIAL: Dependencia funcional, aleatoria e independencia.

Estos conceptos son los más importantes en el estudio de la correlación y regresión, y es fundamental que los alumnos lleguen a diferenciarlos. Para conseguir este objetivo, presentaremos actividades variadas con distintos grados de dependencia entre las variables.

Debemos hacer observar a los alumnos que en una dependencia funcional a cada valor de una variable X (independiente) corresponde un solo valor de otra variable Y (dependiente), pero en la dependencia aleatoria a cada valor de X corresponde una distribución de valores de Y, por lo que este concepto amplía el de dependencia funcional. Además, es posible medir la intensidad y dirección de la relación lineal mediante el coeficiente de correlación de Pearson y la dependencia funcional puede considerarse un caso extremo de la aleatoria (Gea, et. al. 2015).

La definición de independencia es importante, pues la comprensión de esta idea es base de muchos temas estadísticos posteriores; por ejemplo, en inferencia, la mayor parte de contrastes estadísticos e intervalos de confianza están definidos bajo el supuesto de independencia estadística de los datos muestrales.

Bajo estas consideraciones plantearemos las siguientes actividades dirigidas trabajando en grupos:

Actividad 1: Facilitaremos a cada grupo un muelle helicoidal, una regla graduada y varios pesos. Se les pedirá que anoten el alargamiento inicial del muelle, así como las mediciones sucesivas según las diferentes cargas en la siguiente plantilla:

Cargas sucesivas	Lecturas sucesivas **(L_i)**	Alargamiento **$\Delta l_i = L_i - L_0$**	Coeficiente de proporcionalidad $K_i = F_i/\Delta l_i$	
g	mm	mm	g/mm	9,8 N/m

Se les pedirá:

- Representación gráfica y que a partir de ella, deduzcan la pendiente.

- Cálculo del coeficiente de proporcionalidad del muelle.

- ¿Qué alargamiento sufrirá el muelle si colocamos el doble del peso más grande?

- ¿Tiene sentido medir el alargamiento del muelle si no colgamos peso?

- Discusión sobre el tipo de dependencia (funcional, estadística o independencia).

Actividad 2: Proporcionaremos a cada grupo una cinta métrica y les pediremos que se midan el largo del pie y que anoten en una plantilla esa medición junto con su edad en meses. Les preguntaremos:

- Representación gráfica a partir de las mediciones.

- ¿A qué se asemeja esa disposición de los puntos? ¿Podríamos deducir su pendiente?

- Según nuestra gráfica, ¿Podríamos deducir la longitud de pie que le corresponde a una persona con el triple de nuestra edad? ¿Y a un recién nacido de 0 meses?

- Discusión sobre el tipo de dependencia (funcional, estadística o independencia).

Actividad 3: Repartimos a cada grupo un recorte de prensa deportiva con los resultados de los equipos de la liga femenina de baloncesto (20 equipos) en el que figura el número de partidos ganados, perdidos y empatados. Les preguntamos:

- Representación gráfica a partir de las observaciones: lugar que ocupa el equipo en la clasificación en la liga y número de partidos ganados.

- ¿A qué se asemeja esa disposición de los puntos? ¿Podríamos deducir su pendiente?

- Según nuestra gráfica, ¿Podríamos deducir la posición de un quipo con 0 partidos ganados? ¿Podríamos predecir los partidos ganados de un equipo que hubiera quedado en la posición 25ª?

- Discusión sobre el tipo de dependencia (funcional, estadística o independencia).

Actividad 4: Se les pide a los alumnos que busquen en internet los 20 países con mayor consumo per cápita de chocolate y los 20 países con mayor número de Premios Nobel. A continuación, se les pide que recojan esa información en una tabla de doble entrada y representen esa información gráficamente.

- ¿A qué se asemeja esa representación gráfica? ¿Podríamos deducir la pendiente?

- ¿Tiene sentido hablar de relación entre estas variables?

- Discusión sobre el tipo de dependencia (funcional, estadística o independencia).

Conceptos y procedimientos tratados en la sesión inicial

Los conceptos que se contemplan en esta sesión inicial son: dependencia funcional, dependencia estadística o aleatoria, independencia, correlación, covarianza y coeficiente de correlación lineal.

Los procedimientos asociados a estos conceptos son: representación gráfica de los valores de dos variables como pares ordenados, así como la identificación del tipo de relación (o la inexistencia de ésta) entre variables. Esta relación, o la ausencia de la misma, se podrá determinar a partir de la representación gráfica, a partir de la formulación matemática o bien a través del propio enunciado de la situación problema.

El tipo de resolución que esperamos de los estudiantes se halla en el Anexo I.

4. RAZONES DE SER DEL OBJETO

4.1 Razones a tener en cuenta para la introducción a la correlación

En un primer término, la razón de ser que vamos a tener en cuenta a la hora de introducir la correlación está vinculada a la toma de decisiones, fundamental en estadística y en investigación empírica. En un segundo término, mostraremos la correlación como base para entender otros conceptos y procedimientos estadísticos más avanzados (estimaciones y predicciones).

4.2 Razones de ser históricas de la correlación

Desde la prehistoria hasta nuestros días, el discernimiento sobre la posible relación que puede existir entre dos sucesos ha sido un aspecto importante del conocimiento humano. "Conocer si los sucesos se relacionan y, con qué intensidad lo hacen, facilita a las personas explicar el pasado, controlar el presente y predecir el futuro" (Crocker, 1981).

La mayoría de nuestros estudiantes de Educación Secundaria creen que los estudios sobre correlación y regresión se deben a Karl Pearson, seguramente por el nombre que se le da al coeficiente de correlación lineal (coeficiente de Pearson) (Stanton, 2001). Sin embargo, aunque Pearson realizó un riguroso desarrollo matemático del coeficiente de correlación, fue el ingenio y la imaginación de Sir Francis Galton lo que le permitió llegar a los conceptos de correlación y regresión tal y como hoy los entendemos. Ello nos da una importantísima lección didáctica: personas con formación inicial diferente a la matemática, como Galton, partiendo de problemas reales y con una fuerte motivación, llegan a descubrir conceptos matemáticos que han tenido una importancia transcendental. Algo parecido a lo que hoy nos proponen los nuevos currículos, donde se sugiere que partiendo de problemas reales se debe conseguir que nuestros estudiantes descubran conceptos matemáticos importantes (Estepa, Gea, Cañadas y Contreras, 2012). Además, se puede aprovechar el ejemplo del nacimiento de la noción de correlación en los trabajos de Galton (a partir del estudio de la herencia biológica) para seguir las recomendaciones de desarrollo de una idea matemática a través de casos reales.

4.3 Problema que constituya una razón de ser del objeto matemático

Si asumimos que los objetos matemáticos emergen del conjunto de prácticas que el estudiante realiza, como una actividad humana de "resolución de problemas" mediada por el lenguaje y "socialmente compartida", debemos tratar de que el objeto matemático trabajado en el aula adquiera significado personal para el estudiante.

En muchas ocasiones, las nociones estadísticas se presentan a los estudiantes de un modo limitado, con un enfoque particular y encasillado, impidiendo adquirir un razonamiento superior, que permita ser generalizado. Por otra parte, dentro de las situaciones-problema de tipo estadístico que suelen proponer los textos escolares, abundan más la práctica de las técnicas estadísticas que la propia reflexión de los resultados obtenidos. En cualquier caso, el estudiante se limita a aplicar la Estadística como un conjunto de fórmulas (que no suele entender) y que ofrecen conclusiones que, en la mayoría de los casos, tristemente, no saben interpretar. En muchas ocasiones, este es el caso de la correlación y regresión.

Por todo esto, nuestra propuesta se basa en un aprendizaje por proyectos en el que los alumnos recojan sus propios datos para analizar la correlación entre ellos. Al trabajar con proyectos se coloca a los alumnos en la posición de tener que pensar en preguntas como las siguientes (Graham, 1987): ¿Cuál es mi problema? ¿Necesito datos? ¿Cuáles? ¿Cómo puedo obtenerlos? ¿Qué significa este resultado en la práctica? (Batanero y Díaz, 2011).

En concreto, propondremos a los alumnos que a través del hilo conductor de "Los Objetivos de Desarrollo Sostenible" (los ODS son una serie de medidas aprobadas por 193 Estados miembro de Naciones Unidas en 2016 para poner fin a la pobreza, proteger el planeta y garantizar el bienestar global de las personas; www.undp.org) ejerciten las siguientes técnicas y procedimientos estadísticos:

- Recogida de datos mediante consulta de fuentes estadísticas (se propone: www.data.worldbank.org)

- Depuración y tabulación de datos

- Elaboración de tablas de frecuencia; recuento y cálculo de frecuencia

- Elaboración de tablas de doble entrada y cálculo de frecuencias condicionadas y marginales.

- Elaboración de gráficos
- Interpretación de tablas y gráficos
- Elaboración de argumentos y conclusiones a partir del análisis de datos
- Estudio de asociación entre variables
- Uso de hojas de cálculo

De los 17 Objetivos de Desarrollo Sostenible (ODS del Programa de Naciones Unidas para el Desarrollo) proponemos a nuestros alumnos trabajar en las siguientes áreas (escogemos estas tres por representar un área económica, un área social y un área medio ambiental):

- Pobreza (ODS 1)
- Igualdad de género (ODS 5)
- Acción por el clima (ODS 13)

El objetivo de este proyecto consiste en mostrar la utilidad de la estadística en el estudio de interrelaciones entre variables para obtener conclusiones fundamentadas matemáticamente.

¿Qué variables están relacionadas con cada una de las áreas? ¿En cuál es la relación directa o inversa? ¿En qué casos la relación podría ser debida a otras variables? ¿Podríamos en alguno de los casos hallar una función matemática para predecir una variable a partir de la otra? ¿Qué tipo de función?

4.4 Metodología a seguir en su implementación en el aula

Pensamos que la mejor forma de ayudar al estudiante a desarrollar su sentido estadístico es basar las clases de estadística en el trabajo con proyectos, bien planteados por el profesor o escogidos libremente por los alumnos. En lugar de introducir los conceptos y técnicas descontextualizadas, o aplicadas únicamente a problemas tipo, difíciles de encontrar en la vida real, se trata de presentar las diferentes fases de una investigación estadística: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado (Batanero, 2013).

Tal y como señalan Anderson y Loynes (1987), la estadística es inseparable de sus aplicaciones, y su justificación final es su utilidad en la resolución de problemas externos a la propia estadística. La historia de la estadística muestra también como ésta recibe ideas y aportes desde áreas muy diversas, donde, al tratar de resolver problemas diversos (transmisión de caracteres hereditarios, medida de la inteligencia, etc.) se han creado conceptos y métodos estadísticos de uso general (correlación, análisis factorial).

Paralelamente al trabajo por proyectos, presentaremos y reforzaremos las nociones necesarias para llevar a cabo el proyecto y que forman parte de la unidad didáctica, en concreto el estudio y cálculo de la correlación, a través de ejercicios prácticos con GeoGebra.

5. SOBRE EL CAMPO DE PROBLEMAS

5.1 Diseño de los campos de problemas

Para el estudio de la correlación estadística proponemos 4 campos de problemas, de modo que permitan trabajar y comprender los diferentes aspectos que conlleva la correlación. Estos campos son:

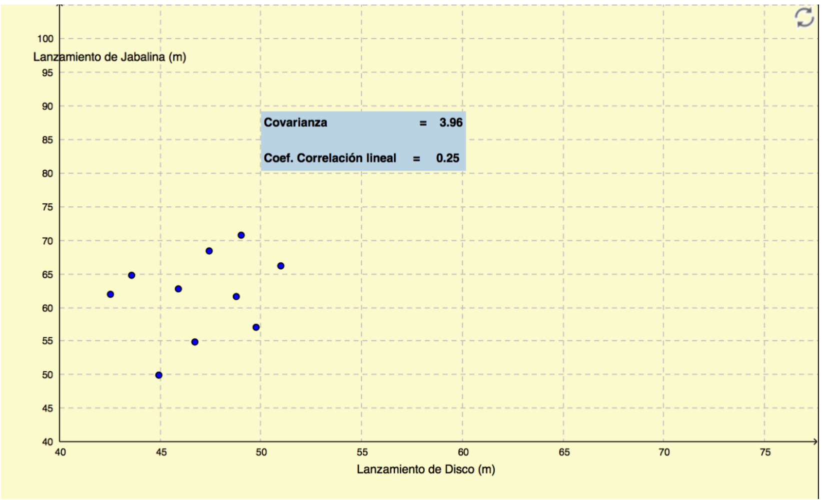
Campo de problemas 1. Aproximación a la correlación.

5.1.1 Representación a través de enunciado, tabla y gráfico.

Problema 1. En la hoja de cálculo se representan los resultados de las 10 primeras clasificadas en la prueba de Heptalon durante los JJOO de Río de Janeiro 2016 en las disciplinas de lanzamiento de peso y lanzamiento de jabalina. En la vista gráfica se representan dichos datos mediante un diagrama de dispersión.

		Javalina (m)	Puntos	Peso (m)	Puntos
LAT	Laura Ikauniece	55,93	975	13,52	762
GRE	Sofia Yfantidou	54,57	949	12,97	725
BEL	Nafissatou Thiam	53,13	921	14,91	855
NED	Nadine Broersen	50,8	876	14,04	797
HUN	Xenia Krizsan	49,78	856	13,78	779
CUB	Yorgelis Rodriguez	48,89	839	13,69	773
FRA	Antoinette Nana Djimou	48,76	836	14,88	853
NED	Anouk Vetter	48,42	830	14,78	846
HUN	Gyorgyi Zsivoczky	48,07	823	14,39	820
GER	Carolin Schafer	47,99	821	14,57	832
CAN	Brianne Thiesen Eaton	47,36	809	13,45	757
GER	Jennifer Oeser	47,22	806	14,28	813
COL	Evelis Aguilar	46,9	800	13,6	767
US	Barbara Nwaba	46,85	799	14,81	848
CZE	Eliska Klucinova	46,73	797	14,41	821
AUT	Ivona Dadic	46,08	784	13,43	756
GBR	Jessica Ennis-Hill	46,06	784	13,86	785
BRA	Vanessa Chefer	45,05	764	13,06	731
NED	Nadine Visser	42,48	715	12,84	717
BAR	Akela Jones	42	706	14,09	800
US	Kendell Williams	40,93	685	11,21	609
US	Heather Miller-Koch	40,25	672	12,91	721
PUR	Alysbeth Felix	40,17	671	11,36	619
GER	Claudia Rath	39,39	656	12,83	716
UKR	Hanna Kasyanova	38,1	631	13,25	744
CZE	Katerina Cachova	37,7	625	12,38	686
GB	Katarina Johnson-Thompson	36,36	598	11,68	640
UKR	Alina Fodorova	35,44	580	14,38	819
NGA	Uhunoma Osazuwa	33,42	542	13,15	737

1.1. Mueve los puntos de la vista gráfica y encuentra la relación existente entre la covarianza y el coeficiente de correlación lineal.

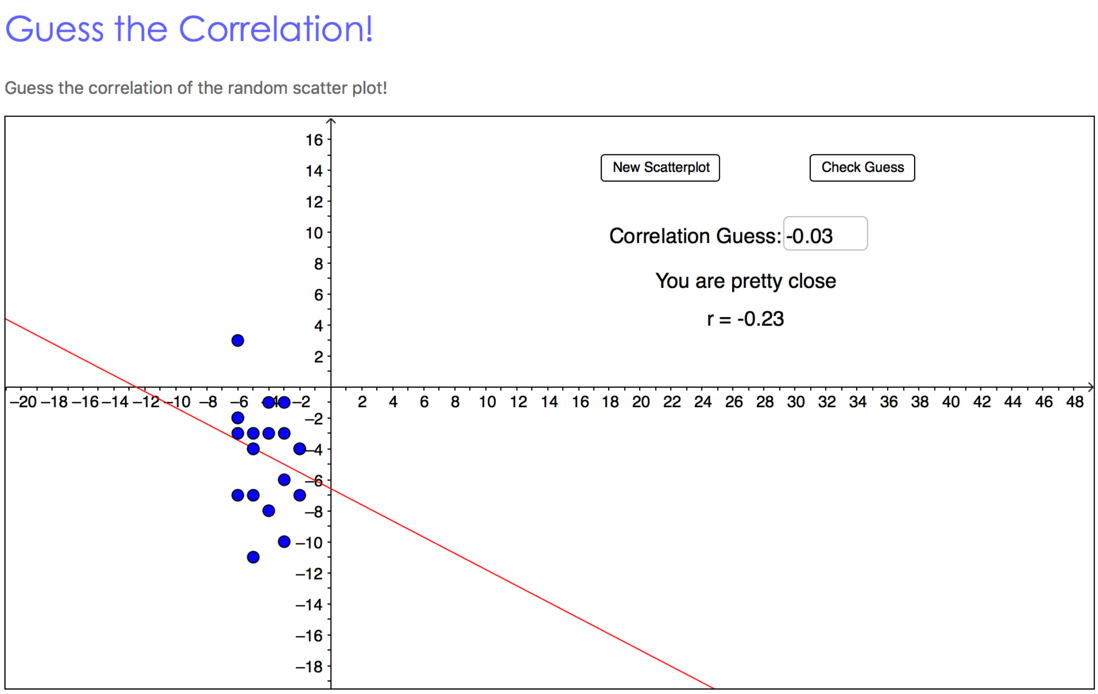


1.2. ¿Qué valores puede tomar el coeficiente de correlación lineal? ¿Y la covarianza?

1.3. Halla el valor de la covarianza si la distancia estuviera expresada en yardas.

5.1.2 Representación gráfica y ecuación.

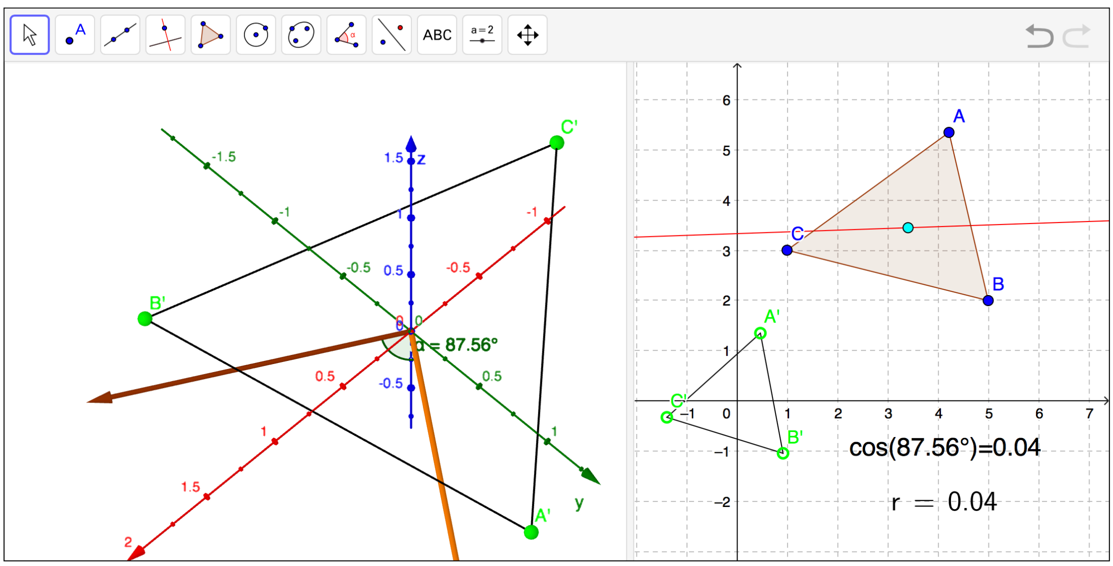
Problema 2. Predice el valor de la correlación según la nube de puntos representada y comprueba cómo varía si movemos alguno de los puntos. ¿Cómo afecta a la correlación un dato muy alejado del centro de gravedad? ¿Cómo es la correlación si la nube de puntos está agrupada de forma aproximada a una disposición vertical?



Campo de problemas 2. Propiedades del coeficiente de correlación

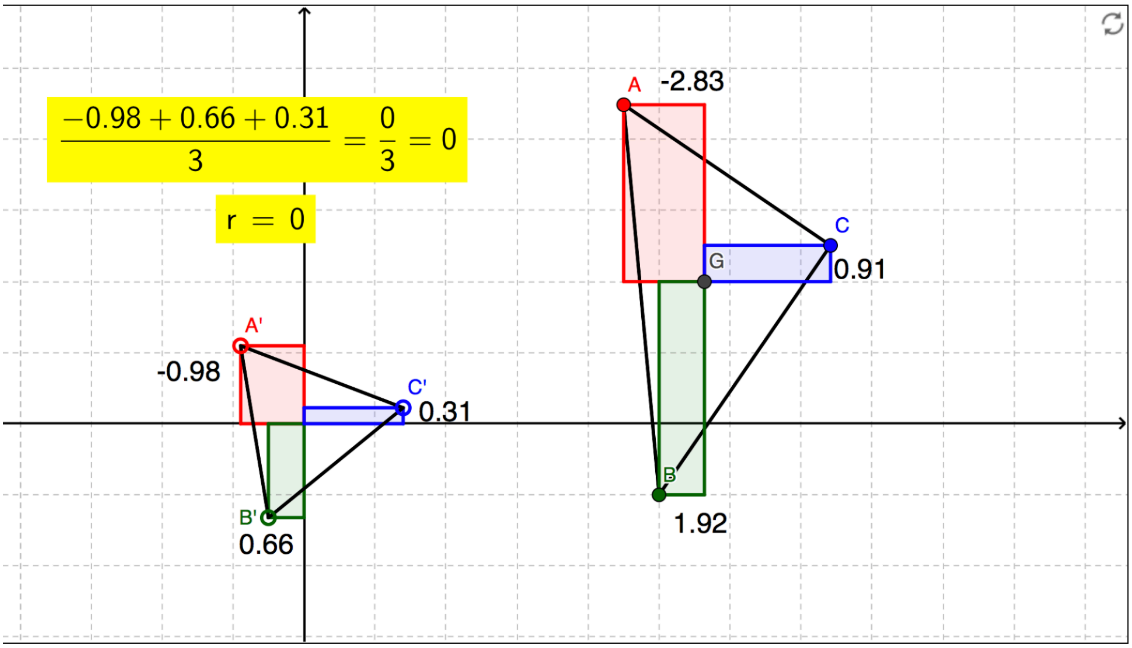
5.2.1 Interpretación geométrica del coeficiente de correlación

Problema 3. Arrastra con cuidado el vértice A del siguiente triángulo ¿Qué relación encuentras entre ese ángulo y el coeficiente de correlación? ¿A qué se debe?



5.2.2 Correlación como la media de las áreas rectangulares

Problema 4. Los puntos A, B y C son los vértices de un triángulo. Forma rectángulos usando cada vértice del triángulo y el baricentro como vértices opuestos y los lados paralelos a los ejes de coordenadas. La estandarización de las coordenadas traslada el baricentro al origen. ¿Cómo es el triángulo estandarizado cuando la correlación es cero? ¿Qué sucede en el promedio de las áreas de los rectángulos cuando la correlación es cero?



Campo de problemas 3. La correlación para tomar decisiones

Problema 5. ¿Crees que la distancia en los lanzamientos está correlacionada con la puntuación obtenida por las atletas? Compruébalo y compara el resultado en los dos tipos de lanzamientos. ¿Cómo influirá en el entrenamiento de las atletas conocer esta correlación?

Campo de problemas 4. Interpretación de la correlación.

Estudio de la correlación en el proyecto de los 3 ODS

Problema 6. Para una investigación estadística, en primer lugar, hay que seleccionar las variables que queremos estudiar para posteriormente recoger los datos.

A través de fuentes estadísticas on line (proponemos el banco de datos del World Bank: https://datos.bancomundial.org/indicador) confecciona las tablas estadísticas para los 217 países con los datos de los últimos 10 años que vayas a utilizar en tu investigación para responder a las siguientes preguntas y realiza las siguientes actividades que se proponen:

- Todos los grupos: Leed los siguientes artículos: "Fin de la pobreza: Por qué es importante" (un.org), "Los beneficios económicos de la igualdad de género" (pactomundial.org), "El clima de 2017, el más caro de la historia" (https://news.un.org).

- Responde a las siguientes preguntas antes de realizar la investigación:

Grupo de la "Pobreza" ODS 1: ¿Cómo se relaciona la educación con la pobreza de los países? ¿Hay diferencia en esa relación si comparamos la tasa de alfabetización en mujeres jóvenes con la pobreza? ¿Cómo crees que será la relación entre el número de niñas y niños trabajando y el PIB?

Grupo de la "Igualdad de género" ODS 5: ¿Crees que cuando el número de mujeres ocupadas aumenta, las economías crecen? ¿Cómo están relacionadas la educación en mujeres y niñas con el crecimiento económico? ¿Cómo influye en el PIB el número de mujeres empresarias?

Grupo de la "Acción por el clima" ODS 13: ¿Crees que la esperanza de vida al nacer está más relacionada con el PIB o con las emisiones de CO2? ¿Cuál crees que será la relación entre las emisiones de CO2 y el PIB? ¿Crees que en los países con mayor número de mujeres empresarias se contamina menos?

- Si no habéis identificado qué variables podemos estudiar para responder estas cuestiones, presentamos la siguiente batería de variables: crecimiento de la población anual, tasa de mortalidad, uso de energía, niños y niñas (entre los 7 y 14 años de edad) trabajando, PIB per cápita, gasto público en educación, tasa de alfabetización entre los jóvenes (15 y 24 años), nivel de estudios terminados por mujeres, prevalencia del VIH, esperanza de vida al nacer, desempleo femenino, emisiones de CO2, uso de energías renovables, exportaciones de bienes y servicios, número de mujeres ocupadas, empresas con participación de mujeres en la propiedad.

- Después de vuestra investigación, responded a las preguntas y comparad las respuestas.

Adjuntamos los contenidos que engloba este proyecto, así como los posibles planteamientos y dificultades en el Anexo II.

5.2 Modificaciones de la técnica inicial

Si bien en los libros de texto tienden a una presentación de la técnica orientada a la aplicación automática de la fórmula de la correlación, el uso de métodos como la representación del diagrama de dispersión (Estepa, 2008), el ajuste gráfico del modelo estadístico junto con el cálculo y la representación de la recta de ajuste mínimo cuadrática son, entre otros, métodos y técnicas de gran utilidad en el Análisis Estadístico Descriptivo.

La mayoría de los problemas propuestos en los libros de texto vienen caracterizados por la homogeneidad, por el predominio de conocimientos procedimentales frente a los conceptuales, la presentación de la técnica orientada a la aplicación automática de la fórmula y el escaso interés por la correcta caracterización de los diferentes conceptos.

En concreto, a través de los campos de problemas propuestos se pretende:

- Dar apoyo sustancial a la actividad investigadora, contra la tendencia de la mayor parte de las transposiciones didácticas de reducir el conocimiento a la técnica,

- Preparar el conocimiento analítico posterior. Si las nociones estadísticas se presentan a los alumnos de un modo limitado, les impide adquirir un razonamiento superior, que permita ser generalizado y no se limite a la práctica particular. Pueden propiciar la impresión de que las técnicas estadísticas son más importantes que los resultados obtenidos. En cualquier caso, el alumno se limita a aplicar la estadística como un conjunto de fórmulas (que en la mayoría de las ocasiones no entiende) y que ofrecen resultados sin ningún contexto (C. Batanero, A. Estepa y J. D. Godino "Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria". Revista Suma, nº 9, 1991. pp 25-31)

5.3 Metodología

Campos de problemas 1 y 2 (aproximación a la correlación y propiedades del coeficiente de correlación respectivamente).

Proponemos el uso de recursos tecnológicos, en concreto Geogebra, para trabajar estos campos de problemas.

El uso de la tecnología en la enseñanza de la estadística ha sido reconocido, entre otros, por Pratt, Davies y Connor (2011), que destacan la reducción del tiempo de cálculo y la ampliación del tipo de gráficos que el alumno puede realizar interactivamente. Igualmente, señalan la posibilidad de trabajar con proyectos, en que el alumno parte de un problema de investigación y, utilizando conjuntos de datos reales, que hoy día son accesibles desde muchas instituciones en Internet, completa todos los pasos de una investigación (Wild y Pfannkuch, 1999).

Una gran ventaja al utilizar estos datos es que se potencia la interdisciplinariedad en clase de estadística, permitiendo aprender contenidos que no se adquieren habitualmente con problemas tomados de los libros de texto; por ejemplo, el efecto de valores atípicos sobre el cálculo de un estadístico (Hall, 2011). Al facilitar el cálculo y la representación gráfica, la tecnología disminuye el problema (tradicional en la enseñanza de la estadística) en cuanto al desfase entre la comprensión de los conceptos y los medios técnicos de cálculo para poder aplicarlos (Batanero y Díaz, 2011).

Campo de problemas 3 (correlación en la toma de decisiones)

Ya hemos indicado que la probabilidad y la estadística son muy cercanas al mundo familiar al alumno y que, por tanto, proporcionan una oportunidad extraordinaria de "matematizar", de mostrar al alumno el proceso de construcción de modelos, así como la diferencia entre "modelo" y realidad. Por otro lado, las teorías de aprendizaje aceptadas con mayor generalidad enfatizan el papel de la resolución de problemas, de la actividad del alumno en la construcción del conocimiento, así como de la formulación (lenguaje matemático), validación (demostración y razonamiento de las ideas matemáticas) e institucionalización (puesta en común; acuerdo social en la construcción del conocimiento).

Los nuevos conocimientos que los alumnos adquieran se apoyarán en los ya conseguidos. Tal y como señala la Orden ECD/489/2016, de 26 de mayo, por la que se aprueba el currículo de ESO y se autoriza su aplicación en los centros docentes de la Comunidad de Aragón: *"los contextos deben ser elegidos para que el alumnado se aproxime al conocimiento de forma intuitiva mediante situaciones cercanas al mismo y vaya adquiriendo cada vez mayor complejidad, ampliando progresivamente la aplicación a problemas relacionados con fenómenos naturales y sociales y a otros contextos menos cercanos a su realidad inmediata."*

Así pues, trabajaremos la resolución de problemas de forma ordenada y progresiva; exponiendo los diferentes procesos mentales para resolver el problema planteado, la formulación de las preguntas correspondientes, los razonamientos y dudas que puedan surgir junto con los errores que se pudieran cometer.

Campo de problemas 4 (interpretación)

Como ya hemos mencionado, los proyectos estadísticos y la experimentación con fenómenos aleatorios juegan un papel primordial en el aprendizaje de la estadística y en concreto, de la correlación y regresión. Los proyectos permiten a los alumnos elegir un tema de su interés en el cual precisen definir los objetivos, elegir los instrumentos de recogida de datos, seleccionar las muestras, recoger, codificar, analizar e interpretar los datos para dar respuesta a las preguntas planteadas. Los proyectos introducen a los alumnos en la investigación, les permiten apreciar la dificultad e importancia del trabajo del estadístico y les hace interesarse por la estadística como medio de abordar problemas variados de la vida real.

6. SOBRE LAS TÉCNICAS

6.1 Sesión inicial

El objetivo de la secuenciación de las cuatro actividades propuestas para esta sesión inicial (apartado 3.3 de este trabajo), es presentar una serie de ejemplos y contraejemplos que sirvan para llevar al alumno a una idea intuitiva conforme la diferencia entre correlación y causalidad.

Si bien es frecuente encontrar en los textos este tipo de ejemplos y contraejemplos como argumento para validar una determinada propiedad, nuestro propósito es generar interés por el objeto matemático a enseñar, así como ayudar a desarrollar el pensamiento inductivo de los alumnos.

Tal y como señalan Crespo y Farfán (2005) en la mayoría de las ciencias, y en particular en la mayoría de los campos de la matemática, se parte de la inducción como método para enunciar sus proposiciones. Sin embargo, son pocos los textos que finalizan el proceso de generalización, limitándose en la mayoría de los casos a dar por “validada” la propiedad o solución.

Las técnicas que se ponen en manifiesto en esta sesión inicial son:

TC 0.1: Representación gráfica a partir de mediciones.

TC 0.2: Cálculo del coeficiente de proporcionalidad del experimento.

TC 0.3: Construcción de una tabla de frecuencias a partir de mediciones.

TC 0.4: Deducción de la pendiente de la recta de regresión a partir de una nube de puntos.

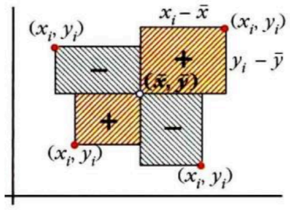
TC 0.5: Estimación del valor de una variable a partir de la nube de puntos de su distribución bidimensional.

TC 0.6: Discusión sobre el tipo de dependencia a partir de una nube de puntos.

6.2 Campos de problemas 1 y 2 (aproximación a la correlación y propiedades del coeficiente de correlación, respectivamente).

Las técnicas que se plantean en este campo de problemas son el uso de representaciones gráficas y algebraicas para apoyar una argumentación verbal o simbólica.

En la mayoría de los textos, se recurre a la representación de los cuatros cuadrantes propuesta por Holmes (2001), y se suele acompañar de una demostración deductiva informal con poco apoyo algebraico donde se comprueba los signos del producto de las diferencias de cada punto al centro de gravedad.



Covarianza

Se llama **covarianza** al parámetro:

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

Ambas expresiones, como es lógico, coinciden. La segunda de ellas es más cómoda para obtener numéricamente la covarianza.

En la figura adjunta, cada sumando $(x_i - \bar{x})\,(y_i - \bar{y})$ de la covarianza es el área de un rectángulo como los que aparecen en la figura.

Según donde esté situado (x_i, y_i) respecto a $(\bar{x}, \bar{y})$, el área $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ será positiva (rojo) o negativa (gris). Si los puntos están próximos a una recta de pendiente positiva, los sumandos son casi todos positivos y la covarianza es grande.

Figura 4. Definición e interpretación de la covarianza (Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). Matemáticas aplicadas a las Ciencias Sociales I. Madrid: Anaya. p. 228)

Las técnicas que se asocian a este campo de problemas 1 "aproximación a la correlación" son:

TC1.1: Dibujo de nube de puntos a partir de tabla de doble entrada.

TC1.2: Cálculo de posibles valores de la covarianza y correlación a partir de una nube de puntos.

TC1.3: Predicción de la correlación según una nube de puntos.

6.3 Campo de problemas 3 (toma de decisiones)

Junto con los problemas mostrados en la sección 5, añadimos las siguientes actividades:

Actividad 1. Las previsiones de la AEMET para una semana en Zaragoza han sido:

mar 29		mié 30				jue 31		vie 01		sáb 02	dom 03	lun 04
12–18 h	18–24 h	00–06 h	06–12 h	12–18 h	18–24 h	00–12 h	12–24 h	00–12 h	12–24 h			
21°C	17°C	15°C	21°C	20°C	16°C							
Probabilidad de precipitación												
100%	100%	85%	100%	100%	65%	20%	20%	5%	60%	75%	80%	75%
Cota de nieve a nivel de provincia (m)												
Temperatura mínima y máxima (°C)												
14 / 24		15 / 22				14 / 26		16 / 27		15 / 29	16 / 28	16 / 27

a) Dibuja la nube de puntos que relaciona la temperatura máxima con la probabilidad media de precipitación cada día.

b) Calcula la correlación.

c) ¿Existe alguna relación entre la temperatura máxima y la probabilidad media de precipitación? ¿En qué te basas? En caso afirmativo, ¿de qué tipo?

Las técnicas que se presentan en esta actividad son: dibujo de una nube de puntos a partir de los datos de una tabla de doble entrada, uso de la calculadora para hallar el coeficiente de correlación e interpretación del coeficiente de correlación a partir de su valor y la representación gráfica de la nube de puntos.

Actividad 2. Proponemos jugar en clase al "Dilema del Prisionero" (Wang et al., 2017) contando con 3 posibles estrategias en cada ronda del juego:

Estrategia 1. Cooperación. Cada participante paga una unidad para que el oponente reciba dos unidades.

Estrategia 2. Deserción. El participante gana una unidad a costa de que el oponente pierda una unidad.

Estrategia 3. Castigo. Se paga una unidad a costa de que el oponente pierda 4 unidades.

Los alumnos participan en varias rondas con dos modalidades: participación anónima (el jugador no conoce a su oponente) y participación descubierta (el jugador sí conoce a su oponente)

Los alumnos apuntarán las fichas ganadas y perdidas junto con la estrategia utilizada, confeccionando ellos mismos su propia tabla de frecuencias.

Analizarán en qué casos la cooperación correlaciona de forma positiva con la ganancia total (será en los casos donde se conoce al oponente, "winners play nice") y en qué modalidad el castigo correlaciona negativamente con la ganancia total (los resultados de la investigación apuntan a que en ambas modalidades el castigo correlaciona negativamente con la ganancia, "winners don't punish").

Las técnicas que se desarrollan en esta actividad son: elaboración de la tabla de valores a partir de un experimento, representación gráfica de una tabla de valores, estimación de la correlación desde el diagrama de dispersión, uso de la calculadora para hallar el coeficiente de correlación e interpretación del coeficiente de correlación a partir de su valor y la representación gráfica de la nube de puntos.

7. SOBRE LAS TECNOLOGÍAS

La técnica de cálculo de la correlación es ya conocida por los alumnos y será el punto de partida de la presente memoria.

Se recordará a los estudiantes que este coeficiente puede interpretarse como la proporción de varianza de la variable dependiente que es explicada por el modelo de regresión. En este sentido, daremos relevancia a la definición estructural de los conceptos, prestando especial atención al significado y a los aspectos interpretativos.

Paralelamente, mediante un razonamiento geométrico basado en la trigonometría, pondremos en relieve la relación entre la correlación con el coseno del ángulo que forman dos variables.

Para justificar la técnica formal de cálculo del coeficiente de correlación también nos basaremos en la demostración algebraica que relaciona el coeficiente de correlación lineal y los coeficientes de regresión a través de una cadena de igualdades:

$$\frac{\sigma_{xy}}{\sigma_x^2} \cdot \frac{\sigma_{xy}}{\sigma_y^2} = \left(\frac{\sigma_{xy}}{\sigma_x \sigma_y} \right)^2 = r^2$$

Todas estas justificaciones están basadas en algunos de los significados históricos del concepto de correlación, los cuales están recogidos en el apartado 4.2 de la presente memoria.

7.1 Proceso de institucionalización

Aunque la responsabilidad de justificar las técnicas las asuma el profesor en todos los campos de problemas junto con la sesión inicial, el proceso de interacción que emerja en cada caso sí será diferente y dependerá del campo de problemas y de los contenidos a tratar. Así, en la “sesión inicial” en la que nuestro objetivo es llevar a los alumnos a una idea intuitiva conforme la diferencia entre correlación y causalidad, facilitaremos interacciones entre los alumnos en el seno de cada equipo y con la clase en su conjunto, para que en último término sea el profesor el encargado de institucionalizar la definición y propiedades de los conceptos explorados.

En el campo de problemas 1 y 2, “aproximación a la correlación” y “propiedades de la correlación” respectivamente, los alumnos presentarán en clase sus soluciones y justificaciones, las cuales se completarán hasta una institucionalización guiada por el profesor.

Será en el campo de problemas 3 y 4, “correlación para la toma de decisiones” e “interpretación”, cuando el profesor sistematice los conocimientos pretendidos mediante los principios socio-constructivistas-interaccionistas.

8. SECUENCIA DIDÁCTICA Y CRONOGRAMA

Planteamos una secuencia didáctica de 8 sesiones de cincuenta minutos de duración cada una, en la que se trabajarán las actividades propuestas en los campos de problemas junto con las dos actividades señaladas en el punto 6.3.

La temporalización de las actividades (Anexo III) es flexible, ya que la aparición de dudas que dificultan el seguimiento de la clase por parte de los alumnos debe ser tenida en cuenta.

Sesión 1. Sesión inicial.

Justificación de la sesión inicial

El sentido de la correlación requiere unas buenas actitudes hacia la estadística y en concreto, hacia el estudio de la correlación y regresión. Debemos lograr que los alumnos se interesen por este tema y su aprendizaje, además de que valoren su importancia para la modelización matemática en diversas áreas de conocimiento (por ejemplo: tecnología/física, social, biológica y deportiva). Mediante esta serie de actividades a través de datos reales fomentamos una actitud crítica hacia el uso inadecuado de la estadística y el interés por detectar sesgos e informaciones erróneas sobre estadística que con frecuencia aparecen en los medios de comunicación.

En esta sesión explicaremos en qué va a consistir el trabajo de investigación sobre los ODS y agruparemos a los alumnos para que analicen las lecturas.

Sesión 2. Campo de problemas 1. Aproximación a la correlación

En el aula de informática con Geogebra. Aproximación deductiva a la correlación. Cómo afectan los datos atípicos y el cambio de unidades.

Sesión 3. Campo de problemas 2. Propiedades del coeficiente de correlación.

En el aula de informática con Geogebra. Se institucionaliza la fórmula de la correlación.

Sesión 4. Campo de problemas 3 y 4. La correlación para tomar decisiones e interpretación del coeficiente de correlación.

Actividades: ejercicio de meteorología y juego del prisionero.

Sesión 5. Campo de problemas 4. Interpretación.

Se termina en clase el proyecto de investigación.

Sesión 6. Exposiciones y repaso.

Sesión 7. Prueba de evaluación cooperativa.

Sesión 8. Revisión de la prueba por grupos.

9. EVALUACIÓN

La prueba constará de un ejercicio sobre lo planteado en la secuencia didáctica sobre el estudio de la correlación estadística.

Para atender a la diversidad presente en el aula, se proponen diferentes apartados: aproximación a la fórmula algebraica, razonamiento de hipótesis, representación gráfica e interpretación de los resultados según el contexto. También se plantean diferentes tipologías de preguntas: tipo test, tipo verdadero y falso, preguntas abiertas, preguntas cerradas. Así mismo, los diferentes apartados estarán secuenciados según la taxonomía de Bloom (recoger información, confirmar la aplicación, hacer uso del conocimiento, desglosar la información, relacionar los resultados y argumentar la evidencia).

La duración de la prueba será de una hora y se permitirá el uso de calculadora. Los alumnos realizarán la prueba de evaluación por grupos (los mismos con los que han trabajado el proyecto de investigación sobre los Objetivos de Desarrollo Sostenible) en la prueba de evaluación continua. En la convocatoria de junio y de septiembre será individual. Todas estas instrucciones, así como el criterio de evaluación será comunicado a los alumnos y a sus familias a principio del curso.

9.1 Diseño de la prueba escrita

El enunciado del problema de la prueba de evaluación es el siguiente:

En nuestra investigación sobre los Objetivos de Desarrollo Sostenible nos encontramos con la siguiente información referida al año 2017:

País	PIB (millones dólares)	País	PIB (millones dólares)
Estados Unidos	14.586.736	Irán	331.015
Kirguistán	4.616	Tailandia	318.522
Guinea	4.511	Chipre	23.132
China	5.815.501	Tanzania	23.057
Japón	5.458.836	Rusia	1.479.929
Alemania	3.391.641	España	1.407.405
Sudáfrica	363.704	Australia	1.131.623

País	Tasa mortalidad infantil	País	Tasa mortalidad infantil
Estados Unidos	5’8	Irán	37’1
Kirguistán	18’8	Tailandia	9’4
Guinea	67’2	Chipre	8’1
China	12’2	Tanzania	41’2

Japón	2	Rusia	6’9
Alemania	3’4	España	3’3
Sudáfrica	32	Australia	4’3

Responded a las siguientes cuestiones indicando los cálculos (con fórmula) y razonamientos:

a) A la vista de los datos de la tabla, ¿creéis que hay alguna relación entre el PIB de un país y su tasa de mortalidad infantil? ¿por qué? (1 punto)

b) Representad gráficamente la dispersión de las dos variables (PIB y tasa de mortalidad infantil) de cada país. (1 punto)

c) ¿Qué opción se ajusta mejor al gráfico dibujado y por qué?: vemos una correlación lineal directa débil; vemos una correlación inversa fuerte; vemos que no hay correlación; vemos una correlación parabólica, vemos una correlación lineal fuerte. (1 punto)

d) Analizad la posible correlación a través del cálculo de la covarianza entre ambas variables. (1 punto)

e) La siguiente frase: “Cuanto mayor es el valor de covarianza, mayor es la correlación entre las variables” ¿es verdadera o falsa? ¿por qué? (1 punto)

f) Calculad el coeficiente de correlación lineal entre el PIB y la tasa de mortalidad infantil. (1 punto)

g) ¿Cómo interpretáis el valor del coeficiente de correlación lineal calculado en el apartado anterior? Si hubiésemos obtenido el mismo valor, pero de signo contrario, ¿qué indicaría? (2 puntos)

h) ¿Podríamos estimar la tasa de mortalidad infantil de un país con un PIB de 0 dólares? ¿y de un país con un PIB de 100.00 dólares? Justificad ambas preguntas y en caso de que se pueda, calculad dicha estimación. (2 puntos).

i) Por último, puntúa a cada compañero/a y a ti mismo/a con un porcentaje (del 0% si no ha aportado absolutamente nada en el examen, hasta el 100% si lo ha dado todo).

Las posibles soluciones de la prueba de evaluación se adjuntan en el Anexo IV.

9.2 Contenidos evaluables

El campo de problemas que evalúa este ejercicio es un cambio entre sistemas de representación: desde el enunciado donde se proponen datos en un contexto, hasta representación gráfica, tabla y ecuación (trabajado desde la sesión inicial hasta la sesión 4, así como durante el proyecto de investigación)

Las técnicas que se ponen en manifiesto son: dibujo de una nube de puntos a partir de una tabla de datos, uso de la calculadora para hallar el valor de la covarianza y del coeficiente de correlación, así como la interpretación del signo y valor de ambos (trabajado en el campo de problemas 1, 2, 3 y 4).

Tareas principales: representación e interpretación gráfica de la dispersión, interpretación del signo y el valor de la covarianza y la correlación lineal, interpretación de la recta de regresión y elección de las variables.

Tareas auxiliares específicas: cálculos algebraicos de la media, varianza y recta de regresión. Expresiones algebraicas de la media, varianza, desviación típica, covarianza y coeficiente de correlación lineal.

Tareas auxiliares generales: simplificación algebraica y operaciones aritméticas.

Los estándares de aprendizaje aplicados en la evaluación de Matemáticas I dentro del Bloque 5 "Estadística y Probabilidad", cumpliendo ECD ORDEN/494/2016, de 26 de mayo, por la que se aprueba el currículo del Bachillerato y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón:

> Est.MA.5.1.1. Elabora tablas bidimensionales de frecuencias a partir de los datos de un estudio estadístico, con variables discretas y continuas.
>
> Est.MA.5.1.2. Calcula e interpreta los parámetros estadísticos más usuales en variables bidimensionales.
>
> Est.MA.5.1.3. Calcula las distribuciones marginales y diferentes distribuciones condicionadas a partir de una tabla de contingencia, así como sus parámetros (media, varianza y desviación típica).
>
> Est.MA.5.1.5. Usa adecuadamente medios tecnológicos para organizar y analizar datos desde el punto de vista estadístico, calcular parámetros y generar gráficos estadísticos.

Est.MA.5.2.1. Distingue la dependencia funcional de la dependencia estadística y estima si dos variables son o no estadísticamente dependientes mediante la representación de la nube de puntos.

Est.MA.5.2.2. Cuantifica el grado y sentido de la dependencia lineal entre dos variables, mediante el cálculo e interpretación del coeficiente de correlación lineal.

Est.MA.5.2.4. Evalúa la fiabilidad de las predicciones obtenidas a partir de la recta de regresión mediante el coeficiente de determinación lineal.

Est.MA.5.3.1. Describe situaciones relacionadas con la estadística, utilizando un vocabulario adecuado (pp. 14107-14108).

9.3 Respuestas y errores esperados

Esperamos razonamientos e interpretación intuitivos. Se usan teorías previas sobre el contexto para argumentar la correlación (tarea principal), en lugar de considerar los datos disponibles.

Los errores que prevemos son: de tipo algebraico, mal uso de la fórmula, confusión entre parámetros (covarianza y correlación) y de interpretación del enunciado, así como de los resultados (estimación, signo y valor de los parámetros calculados).

Adjuntamos en el Anexo IV el detalle de las posibles respuestas y errores esperados por apartados.

9.4 Criterios de calificación

Modelo de penalización de errores: Modelo de tercios (Gairín, Muñoz y Oller, 2012).

La nota de la prueba de evaluación continua de la propuesta didáctica sobre correlación quedará como sigue:

5 puntos máximo del proyecto de investigación (autoevaluación y coevaluación):

2’5 puntos máximo, la valoración global del trabajo por parte de la profesora.

2’5 puntos máximo, será la valoración de los compañeros y autovaloración (cada uno se puntúa así mismo y a sus compañeros con una nota del 0 al 10 y calculamos una media aritmética respecto a los dos puntos y medio).

5 puntos máximo prueba de evaluación grupal:

Con la nota que califique la profesora cada prueba grupal, se calculará la nota de cada alumno según su porcentaje medio de aportación en el examen (media que se calculará a partir de los porcentajes que han indicado los propios alumnos en el examen). De esta manera, si el examen tiene un 10 y un alumno ha aportado el 100% obtendrá los 5 puntos.

Para asegurarnos de que la evaluación sea formativa, corregiremos el examen en la sesión posterior. Repartiremos los exámenes (sin que conste las correcciones de la profesora) y serán los alumnos (en grupo) los que vayan revisando el examen de otro grupo. Si la nota que otorgan los grupos no difiere en más de medio punto con la puntuación que ha asignado la profesora, la calificación final en esta prueba será la nota más alta de las dos. Si la calificación con la que corrigen los grupos difiere en medio punto o más, la nota final en esta parte será la que otorgue la profesora.

BIBLIOGRAFÍA

Anderson, C. W. y Loynes, R. M. (1987). The teaching of practical statistics. New York: Wiley.

Aparicio, A.M., Arribas, F., González, C., Llorente, J. y Ruiz, M.J. (2015). *Matemáticas I.* Madrid: Editex.

Arteaga, P., Batanero, C., Gea, M.M., López-Martín, M.M. (2017). Análisis de los conceptos asociados a la correlación y regresión en los textos de Bachillerato. *Didacticae, 1*, 60-76.

B. S. Bloom (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Michigan: Longsmans.

Barbancho, A. G. (1973). Estadística elemental moderna. Barcelona: Ariel.

Batanero, C., Estepa, A. (1995). Concepciones iniciales sobre la asociación estadística. *Enseñanza de las ciencias, 13*(2), 155-170.

Batanero, C. (2013). Sentido estadístico: Componentes y desarrollo. En *Jornadas Virtuales de Didáctica de la Estadística, Probabilidad y Combinatoria* (pp.1-6). Granada.

Batanero, C. y Díaz, C. (2011). Estadística con proyectos. Granada.

Contreras, J. M., Batanero, C., Godino, J. D., Cañadas, G.R., Arteaga, P., Molina, E., Gea, M.M. y López, M.M. (2015). Didáctica de la Estadística, Probabilidad y Combinatoria. En *Jornadas Virtuales de Didáctica de la Estadística, Probabilidad y Combinatoria* (pp.101-117). Granada.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin 90*(2), 272-292.

Datos de libre acceso del Banco Mundial. Recuperado de https://datos.bancomundial.org

Estepa, A. (2008). Interpretación de los diagramas de dispersión por estudiantes de Bachillerato. Enseñanza de las Ciencias 26 (2), 257-270.

Estepa, A., Gea, M.M., Cañadas, G.R. y Contreras, J.M. (2012). Algunas notas históricas sobre la correlación y regresión y su uso en el aula. *Números*, *81*, pp. 5-14.

Gairín, J.M., Muñoz, J.M., Oller, A.M. (2012). Propuesta de un modelo para la calificación de exámenes de matemáticas. En *Investigación en educación matemática XVI* (pp. 261-274). Jaén.

Gea, M. M., Batanero, C., Cañadas, G. R. y Contreras, J. M. (2013). Un estudio empírico de las situaciones-problema de correlación y regresión en libros de texto de bachillerato. En: A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds),. Investigación en Educación Matemática XVII (pp. 293-300). Bilbao: Sociedad Española de Investigación en Educación Matemática.

Gea, M.M., Batanero, C., Arteaga, P., Cañadas, G.R., Contreras, J,M. (2014). Análisis del lenguaje sobre la correlación y regresión en libros de texto de bachillerato. *SUMA, 76*, 37-45.

Graham, A. (1987). Statistical investigations in the secondary school. Cambridge: The Open University Centre for Mathematics Education.

Hall, J. (2011). Engaging teachers and students with real data: benefits and challenges. En: C. Batanero, G. Burrill, y C. Reading (eds.). Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI and IASE study (pp. 335-346). Nueva York: Springer.

Holmes, P. (2001). Correlation: From picture to formula. *Teaching Statistics*, *23(3)*, 67-71.

Lavalle, A.L., Micheli, E.B. y Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *Revista Latinoamericana de Matemática Educativa, 9(3),* 383-406.

Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa, BOE núm. 295 § 12886 (2013).

Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp.221-255). Dordrecht: Kluwer.

Objetivos de Desarrollo Sostenible. Programa de las Naciones Unidas para el desarrollo. Recuperado de http://www.undp.org/content/undp/es/home/sustainable-development-goals.html

Orden ECD/489/2016, de 26 de mayo, por la que se aprueba el currículo de la Educación Secundaria Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón

Orden ECD/494/2016, de 26 de mayo, por la que se aprueba el currículo del Bachillerato y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón, BOA núm. 106 (2016).

Ouvrier-Buffet, C. (2004). Construction of Mathematical Definitions: An epistemological and didactical study. In M. J. Hoines & A.B. Fuglestad (Eds.), Proceedings of the 28 th Conference of the International Group for the Psychology of Mathematics Education (Vol. 3, pp. 473-480).

Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato, BOE núm. 3 § 37 (2015).

Reid, A. y Pecotz, P. (2002). Students' Conceptions of Statistics:A Phenomenographic Study. *Journal of Statistics Education*, *10(2)*. Recuperado de http://www.amstat.org/publications/jse/v10n2/reid.html

Sánchez Cobo, F. T., Estepa, A. y Batanero, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias, 18(2)*, 297-310.

Sánchez Cobo, F.T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios.* Tesis doctoral. Universidad de Granada.

Stanton, J.M. (2001). Galton, Pearson, and the peas: a brief history of linear regression for statistics instructors. *Journal of Statistics Education*, *9(3)*. Recuperado de http://www.amstat.org/publications/jse/v9n3/stanton.html

Wang, Z., Jusup, M., Wag, R-W., Shi, L., Iwasa, Y., Moreno, Y. y Kurths, J. (2017). Onymity promotes cooperation in social dilemma experiments. *Science Advances*, *3*, 1-7. Recuperado de http://advances.sciencemag.org/content

Wild, C.J. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. International Statistical Review 67(3), 223-265.

Zieffler, A, y Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. Statistics Education Research Journal, 8 (1), 7- 31.

ANEXO I

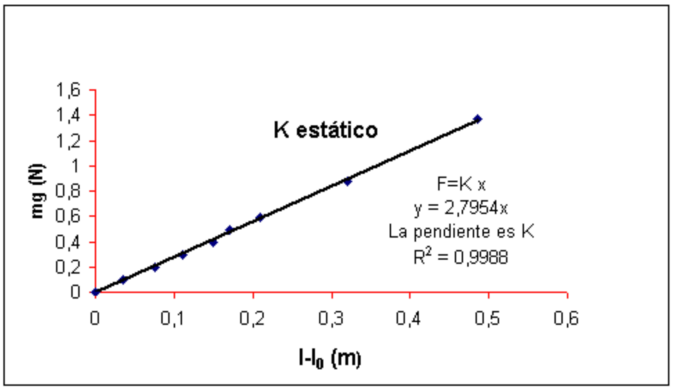
Actividad 1: Facilitaremos a cada grupo un muelle helicoidal, una regla graduada y varios pesos. Se les pedirá que anoten el alargamiento inicial del muelle (L_0), así como las mediciones sucesivas según las diferentes cargas en la siguiente plantilla:

l-lo (m)	mg (N)	m (kg)	K (estática)
0,11	0,294	0,03	2,673
0,15	0,392	0,04	2,613
0,17	0,49	0,05	2,882
0,21	0,588	0,06	2,800
0,32	0,882	0,09	2,756
0,49	1,372	0,14	2,829

Se les pedirá:

- Representación gráfica y que, a partir de ella, deduzcan la pendiente.

Una primera barrera en la resolución de esta cuestión posiblemente sea la determinación de la variable independiente y dependiente. Para conducir a un planteamiento correcto, se propondrá discutir los posibles planteamientos con los compañeros del grupo, siendo este trabajo supervisado por el profesor. Para ello, podría resultar conveniente guiar a los alumnos por medio de las siguientes preguntas: ¿qué datos son los que tenemos que recoger? ¿Qué es lo que tenemos que averiguar?



Para el cálculo de la pendiente se prevén dos caminos alternativos: uno de ellos a través de la aplicación de la unidad didáctica anterior de funciones (hallando la expresión de la recta que pasa por dos puntos o bien calculando la pendiente mediante el vector director) y, otro, a través del cálculo de la constante del muelle (hallando un valor promedio de k=2’8 N/m).

- Cálculo del coeficiente de proporcionalidad del muelle.

En la representación del apartado anterior los alumnos podrán observar que algún valor se desvía de la media, sobre todo, si el valor que se aleja está hacia el límite del alargamiento (lo que puede ser debido a que se ha sobrepasado el límite de elasticidad del muelle).

- ¿Qué alargamiento sufrirá el muelle si colocamos el doble del peso más grande?

Para resolver esta cuestión, los alumnos podrán plantear una resolución de tipo "proporcionalidad directa" (al doble de peso le corresponde el doble de alargamiento) ; o bien, a través de la sustitución de la variable peso en la ecuación de la recta.

- ¿Tiene sentido medir el alargamiento del muelle si no colgamos peso?

Aquí los estudiantes podrán comparar sus razonamientos de tipo deductivo con las conclusiones que puedan arrojar a la vista de la gráfica junto con el apoyo de tipo más analítico. Obviamente, si no se carga ningún peso, el alargamiento del muelle es nulo (de ahí que la recta pase por el origen de coordenadas).

- Discusión sobre el tipo de dependencia (funcional, estadística o independencia).

El propósito de esta actividad es recordar a los alumnos la definición de la dependencia funcional a través de la propia experimentación y la búsqueda de una relación matemática.

Actividad 2: Proporcionaremos a cada grupo una cinta métrica y les pediremos que se midan el largo del pie y que anoten en una plantilla esa medición junto con su edad en meses. Les preguntaremos:

- Representación gráfica a partir de las mediciones.

Al igual que en la actividad anterior, si los alumnos tienen dificultad para decidir qué variable es la independiente y cuál es la dependiente, les preguntaremos ¿Qué variable es la que sabemos y qué variable es la que tenemos que medir?

Una posible nube de puntos puede ser la siguiente:

- ¿A qué se asemeja esa disposición de los puntos? ¿Podríamos deducir su pendiente?

Lo interesante de esta cuestión es que los estudiantes propongan diferentes alternativas de respuesta, aunque la mayoría deduzcan que se asemeja a una recta de pendiente casi nula.

- Según nuestra gráfica, ¿Podríamos deducir la longitud de pie que le corresponde a una persona con el triple de nuestra edad? ¿Y a un recién nacido de 0 meses?

Al comentar la respuesta del apartado anterior entre todos, muchos alumnos se darán cuenta que la edad y la longitud del pie no tienen una dependencia funcional exacta. Así pues, la conclusión a la que esperamos que lleguen nuestros estudiantes es que a una persona con el triple de edad no le corresponde el triple de longitud de pie (al igual que no tiene sentido que a un recién nacido le corresponda una longitud de 0 cm).

- Discusión sobre el tipo de dependencia (funcional, estadística o independencia).

El objetivo de esta actividad es hacer ver, a través de ejemplos experimentales, la diferencia entre dependencia funcional y dependencia estadística o aleatoria. Estos conceptos muchas veces son confundidos por los estudiantes, identificando relación funcional con la existencia de relación no lineal (sin tener en cuenta que puede haber dependencia estadística de tipo no lineal).

Muy similar es el razonamiento, discusión y conclusión de la **actividad 3**.

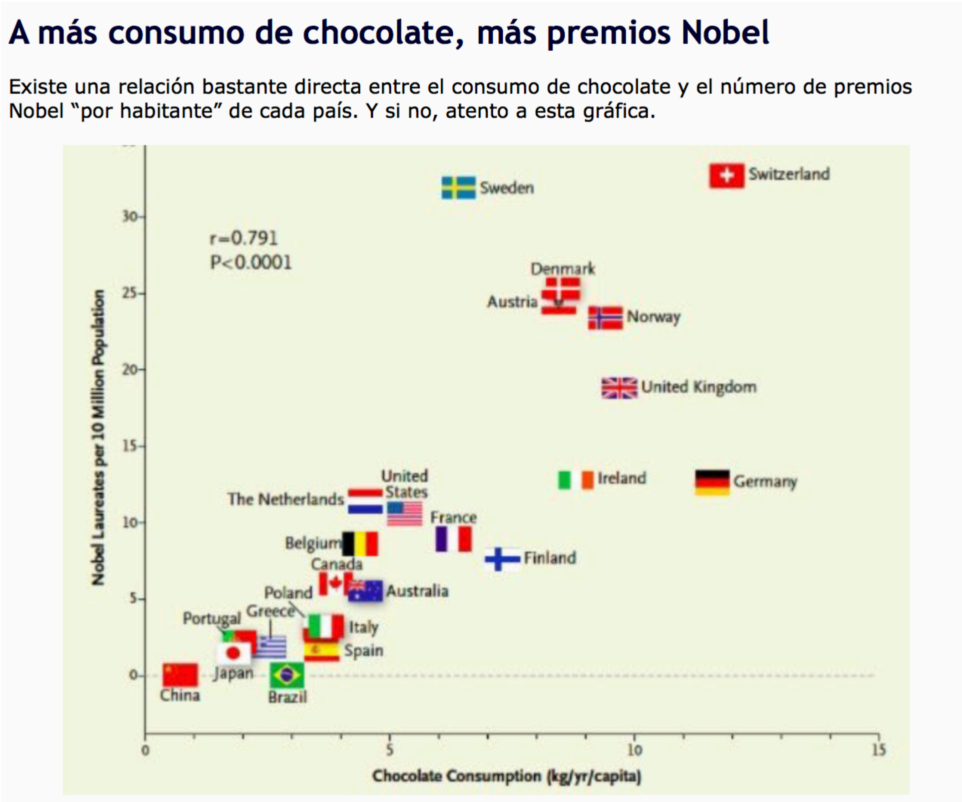
Actividad 4: Se les pide a los alumnos que busquen en internet los 20 países con mayor consumo per cápita de chocolate y los 20 países con mayor número de Premios Nobel. A continuación, se les pide que recojan esa información en una tabla de doble entrada y representen esa información gráficamente.

- ¿A qué se asemeja esa representación gráfica? ¿Podríamos deducir la pendiente?

- ¿Tiene sentido hablar de relación entre estas variables?

- Discusión sobre el tipo de dependencia (funcional, estadística o indepen-dencia).

El objetivo de esta actividad es conseguir que los estudiantes sean críticos con el tipo de información que ofrecen las diversas fuentes que puedan consultar. A modo de ejemplo, un titular que pueden encontrar en internet sería: "A más consumo de chocolate, más premios Nobel".



El objetivo de esta actividad es que los estudiantes lleguen a la conclusión de que la correlación no implica causalidad y que estamos delante de una falacia aunque el coeficiente de correlación nos indique el alto grado de relación entre las variables de estudio.

ANEXO II

Objetos matemáticos que emergen y contenidos que trata este proyecto:

Un primer paso es la discusión de las variables y el problema de medición. Los alumnos se van a encontrar con que las variables que escojan vienen dadas en diferentes escalas de medición o categorizaciones. La toma de conciencia sobre la complejidad del proceso de elaboración de las estadísticas demográficas o económicas es un paso importante para valorar el trabajo estadístico; además, servirá de reflexión sobre las diferentes unidades estadísticas que pueden usarse en un estudio y la necesidad de trabajar con valores aproximados o de combinación para obtener índices globales.

En segundo lugar, al tener que calcular los primeros parámetros, los estudiantes se darán cuenta de la sensibilidad/robustez de ciertos estadísticos ante los valores atípicos (concepto que será recordado por el profesor).

En última instancia, al analizar las relaciones entre dos variables numéricas los alumnos deben extender la idea de dependencia funcional a dependencia aleatoria y diferenciar sus tipos (lineal o no; directa o inversa) así como graduar empíricamente la intensidad de la relación contrastándola con la pre-creencia intuitiva.

Este proyecto también se encamina a la búsqueda de modelos que puedan predecir las variables explicadas en función de las variables explicativas, lo que de nuevo conecta con otro contenido del currículo de matemáticas: las funciones.

Un ejemplo de posibles nubes de puntos que los estudiantes pueden obtener sería el siguiente:

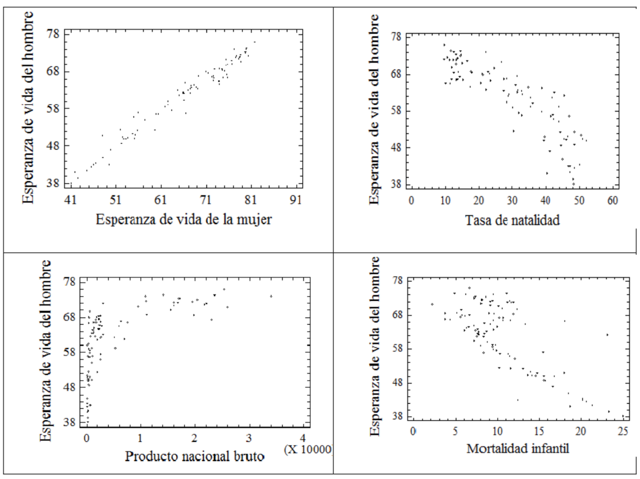


Figura 5. Relación de la esperanza de vida del hombre con otras variables

Calculando los diferentes coeficientes de correlación, los alumnos comprobarán la conveniencia de representar gráficamente los datos en un diagrama de dispersión y podrán observar diferentes tipos de relaciones (relación inversa no lineal, por ejemplo, en el caso de la tasa de natalidad y el PNB).

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	-,629(a)	,396	,389	10,70878

a Variables predictoras: (Constante), Producto nacional bruto en millones de dólares

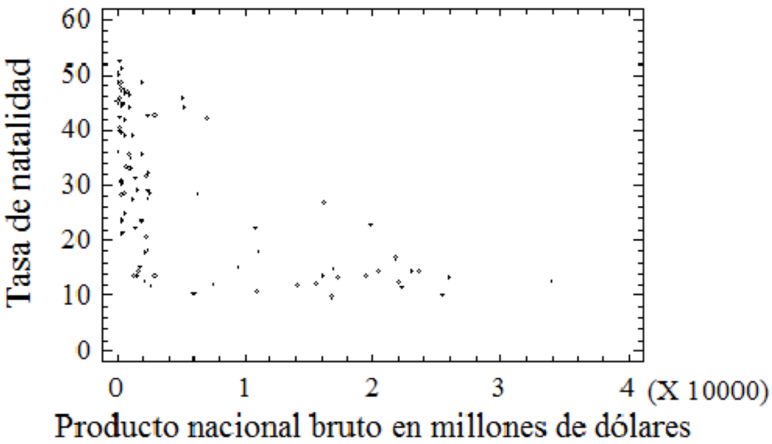


Figura 6. Relación del PNB y la tasa de natalidad

Será en este punto cuando el profesor introduzca la posibilidad de estudiar la relación entre dos variables a través de otros modelos diferentes al lineal.

Algunas dificultades y errores previsibles:

Las dificultades y errores en torno a las nociones de dependencia y correlación pueden estar ligadas a las ideas de condicionamiento y causalidad, puesto que, en general, construimos el conocimiento del mundo sobre la base de relaciones de causa y efecto entre diferentes sucesos. Por ello, la tarea del profesor será institucionalizar que "causalidad implica una asociación, pero no al contrario" (institucionalización que será expuesta en la sesión inicial de esta unidad), puesto que una asociación estadística entre variables puede ser debida a otras variables intervinientes o incluso ser espúrea y no implica relación causal.

Junto con la anterior barrera inicial indicada, señalamos las siguientes dificultades adicionales: distinguir si dos variables presentadas conforman una distribución bidimensional, dificultad de relacionar la magnitud de la covarianza con la intensidad o relacionar los coeficientes de regresión con el signo de la correlación, estimar el coeficiente de correlación desde una tabla con datos en lugar de la usual gráfica (diagrama de dispersión), comparación entre los diferentes valores del coeficiente de correlación (errores en los conocimientos sobre el orden de números enteros) y la transitividad de la correlación (los alumnos pueden pensar que si una variable A está correlacionada con otra B y esta está correlacionada con una variable C, las variables A y C debieran estar correlacionadas, pero en general, la correlación no es transitiva). Esto se debe a una extrapolación de propiedades que se cumplen en otras relaciones (como la de orden).

ANEXO III

Sesión 1: Sesión inicial (en el aula)	Introducción
	Actividades 1, 2, 3 y 4
	Institucionalización de definiciones
	Formación grupos para el proyecto de investigación
Sesión 2: Aproximación a la correlación (aula informática)	Recordamos
	Problema 1: Heptalón en Río'2016
	Problema 2: Guess the correlation
	Institucionalización de la fórmula de la covarianza y desviaciones típicas de las variables marginales
	Tiempo para el trabajo proyecto de investigación
Sesión 3: Propiedades del coeficiente de correlación (aula informática)	Recordamos
	Problema 3: Vértices, ángulos y correlación
	Problema 4: Áreas y correlación
	Institucionalización de la fórmula del coeficiente de correlación
	Tiempo para el trabajo proyecto de investigación
Sesión 4: Toma de decisiones (aula)	Recordamos
	Problema 5: ¿Cómo afecta a nuestro entrenamiento?
	Actividad 1: Meteorología
	Actividad 2: Juego del prisionero

	Tiempo para el trabajo proyecto de investigación
Sesión 5: Interpretación (aula)	Recordamos
	Tiempo para el trabajo proyecto de investigación
Sesión 6: Exposiciones (aula)	Exposiciones
	Repaso
Sesión 7: Prueba conocimientos (aula)	Tiempo de realización de la prueba
	Tiempo para calificar el trabajo cooperativo (autoevaluación y evaluación de compañeros)
Sesión 8: Revisión (aula)	Revisión de la prueba por parte del profesor y seguimiento de los alumnos
	Calificaciones de los alumnos

ANEXO IV

a) A la vista de la tabla ¿Hay algún tipo de relación?

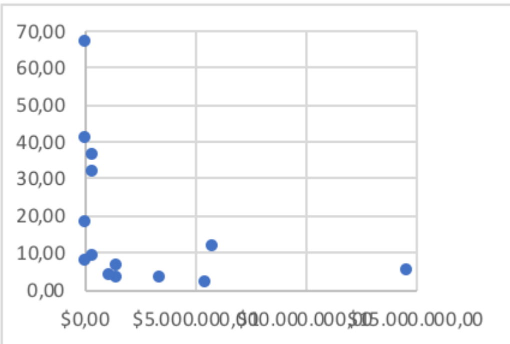
Posiblemente, los alumnos se fijen en los datos más extremos y comparen su relación.

Esta pregunta tiene como objeto establecer un primer análisis exploratorio de datos (detectar datos atípicos y validar la conveniencia o no de la dependencia).

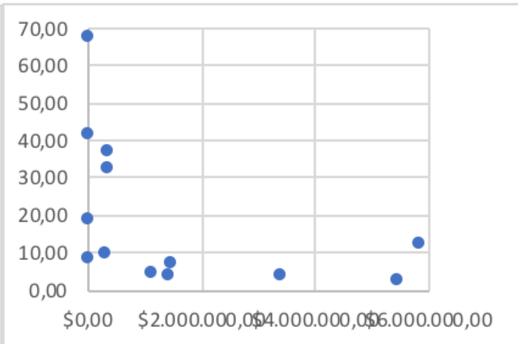
La naturaleza de esta cuestión se ha tratado tanto en la sesión inicial como en una primera fase del proyecto de investigación.

b) Representación gráfica

Prevemos dos posibles representaciones. Puesto que hay un dato atípico (EEUU) habrá grupos que se den cuenta que el valor que arroja este país distorsiona mucho los parámetros que se vayan a calcular.



Con todos los datos



Sin el dato de EEUU

Las dificultades que prevemos en este apartado son: la incorrecta identificación de las variables como explicada y explicativa, así como la dificultad a la hora de trabajar con escalas diferentes y la no identificación inmediata de un posible patrón esperado (una relación lineal directa).

c) ¿Qué opción se ajusta mejor al gráfico dibujado y por qué?: vemos una correlación lineal directa débil; vemos una correlación inversa fuerte; vemos que no hay correlación; vemos una correlación parabólica, vemos una correlación lineal fuerte. (1 punto)

Al igual que en el apartado anterior, tampoco hay una única respuesta correcta. Habrá grupos que opten por indicar una correlación inversa fuerte y otros que se inclinen por la opción de la inexistencia de relación. Ambas respuestas serán consideradas como válidas si sus argumentaciones correspondientes siguen las siguientes líneas:

Si el grupo de estudiantes opta por indicar que hay correlación inversa fuerte, la justificación que la argumente deberá basarse en la pendiente negativa y/o que a mayores valores de la variable PIB le corresponden menores valores de la variable Tasa de Mortalidad Infantil.

Si el grupo se inclina por concluir que no encuentra relación, deberá argumentar que no se encuentra relación lineal pero sí puede existir otro tipo de relación.

La calificación será la siguiente: si no identifican uno de los dos tipos de relación (tarea principal) se penaliza todo el ejercicio (0 puntos). Si identifican uno de los dos tipos pero la argumentación es incorrecta o incompleta se penaliza con la mitad de la puntuación máxima del ejercicio (0'5 puntos).

d) Analizad la posible correlación a través del cálculo de la covarianza entre ambas variables. (1 punto)

COV (X,Y)= -27944597

Las dificultades que prevemos para resolver esta cuestión son tanto de tipo aritméticas como de tipo estadísticas.

Si bien los números son valores muy altos, los alumnos ya han aprendido a través del trabajo de investigación a redondear para aligerar los cálculos. Por eso, y por su frecuente uso en la realidad, creemos en la conveniencia del trabajo con este tipo de valores. Consideraremos los errores de tipo aritmético como errores de tareas auxiliares generales y los penalizaremos descontando hasta 0,3 puntos.

El error en la fórmula, así como la incorrecta interpretación del signo negativo serán considerados como fallos en la tarea principal y serán penalizados hasta 1 punto.

El alto valor de la covarianza viene dado por la propia naturaleza de los datos. Si hay algún grupo que lo asocia a la intensidad de la correlación será penalizada hasta 0,5 puntos.

e) La siguiente frase: "Cuanto mayor es el valor de covarianza, mayor es la correlación entre las variables" ¿es verdadera o falsa? ¿por qué? (1 punto)

En esta cuestión esperamos las siguientes respuestas:

Definiciones de la covarianza y correlación (tarea principal).

Fórmulas de la covarianza y correlación (tarea principal).

Ejemplo (texto y/o gráfico) que argumente la frase enunciada (tarea principal).

Interpretación de la relación entre ambas a través del cociente (tarea principal).

Daremos la puntuación máxima a cualquiera de las respuestas alternativas expuestas correctamente. La penalización por el incorrecto argumento será de hasta el 100%.

f) Calculad el coeficiente de correlación lineal entre el PIB y la tasa de mortalidad infantil. (1 punto)

Coeficiente de correlación: -0,3873209

En este apartado los errores que se prevén son de tipo aritmético (consideradas como tareas auxiliares generales y penalizadas con un -0,3 puntos máximo), así como errores en las expresiones algebraicas de las fórmulas de la correlación, covarianza y desviaciones típicas (tareas principales que conllevarán una penalización máxima si no se expresan correctamente).

g) ¿Cómo interpretáis el valor del coeficiente de correlación lineal calculado en el apartado anterior? Si hubiésemos obtenido el mismo valor, pero de signo contrario, ¿qué indicaría? (2 puntos)

En cuanto a la interpretación del valor del coeficiente de correlación, los estudiantes deberán indicar:

Que se encuentra próximo a cero, con lo que la correlación lineal entre ambas variables es algo débil. Este argumento será considerado como tarea principal. Si los alumnos omiten esta respuesta, la penalización de este apartado será máxima.

Consideramos la posibilidad de que algún grupo observe y argumente el porqué de este valor (ya sea dado por la escasa cantidad de datos a analizar o bien, porque el tipo de relación no sea lineal sino de otro tipo). Aunque contemplamos esta posibilidad de argumentación, no se exige para la calificación de este apartado.

En cuanto a la interpretación del signo negativo, los alumnos deberán responder que se trata de una relación inversa/indirecta/correlación lineal negativa. Esta interpretación del signo es considerada como tarea principal y, por tanto, según el modelo que estamos siguiendo en la calificación, su omisión o error podrá conllevar la total penalización de este apartado.

h) ¿Podríamos estimar la tasa de mortalidad infantil de un país con un PIB de 0 dólares? ¿y de un país con un PIB de 100.00 dólares? Justificad ambas preguntas y en caso de que se pueda, calculad dicha estimación. (2 puntos).

Las respuestas que esperamos son tanto afirmativas como negativas. Ambas las consideraremos correctas si el argumento es acertado.

Los alumnos deberán relacionar esta respuesta con el valor del coeficiente de correlación obtenido en el apartado f). Si ese apartado no lo habían contestado, deberán indicar en su respuesta igualmente la relación de la estimación de una variable con respecto a otra según el valor de este coeficiente. Consideramos este razonamiento como tarea principal (con su penalización correspondiente según el modelo de tercios que seguimos).

Para calcular la estimación que se pide, contemplamos tres posibilidades: utilizando una interpolación, a través de la sustitución de la variable en la recta que pasa por dos puntos o bien, por medio de la recta de regresión. Consideraremos válidos cualquiera de los tres métodos.

Los errores aritméticos en la estimación serán considerados como errores en las tareas auxiliares generales.

Si los estudiantes no logran calcular la estimación pero sí identifican y argumentan el método por el cual lo podrían calcular, penalizaremos con -0,3 puntos sobre el total de la puntuación máxima del ejercicio.