







EVALUACIÓN GENERAL DE DIAGNÓSTICO



etsinf





Índice de la memoria

1.	Introducción
2.	Finalidad y objetivos
3.	Calidad y limpieza de datos
4.	Análisis exploratorio de datos
	4.1. Metodología
	4.2. Análisis descriptivo.
	4.2.1. Descripción de la muestra final y resumen del conjunto de datos
	después del preproceso.
	4.2.2. Análisis unidimensional de las variables a analizar.
	4.2.2.1. Descripción de las variables más representativas.
	4.2.3. Análisis multidimensional de las variables a analizar.
	4.2.3.1. Estandarizar las variables.
	4.2.3.2. Identificación, estudio y descripción de relaciones.
5.	Conclusiones
6.	Anexos
_	



1.INTRODUCCIÓN

Resumen:

La evaluación general de diagnóstico tiene como objetivo contribuir a la mejora de la calidad de la educación a partir del conocimiento de la situación del sistema educativo al que se llega, en este caso, a través de la valoración de los aprendizajes de los estudiantes y de los procesos de innovación y mejora de la educación en todo el sistema. La base de datos correspondiente se elaboró en 2009 con los datos de Cuarto Curso de la Educación Primaria.

Población:

Para la aplicación de las pruebas de evaluación y de los cuestionarios de contexto se han extraído muestras a partir de las poblaciones recogidas en la tabla siguiente:

Poblaciones	Información proporcionada
 Alumnado de 4º curso de Educación Primaria de ámbito estatal. 	 Nivel alcanzado en competencias básicas Variables de contexto y de procesos.
 Profesorado del alumnado de 4º curso de Educación Primaria de ámbito estatal. 	Variables de recursos y de procesos.
 Directoras y directores de centros escolares de ámbito estatal que imparten Educación Primaria. 	Variables de contexto y de recursos.
Familias del alumnado de 4º curso de Educación Primaria de ámbito estatal.	Variables de contexto y de recursos.

Muestra:

A partir de la población del alumnado de cuarto curso de Educación Primaria se ha extraído la muestra.

Para el tamaño de la muestra se fijaron 50 centros por comunidad autónoma. En el territorio administrado por el Ministerio de Educación, al tener menos centros de los fijados, la muestra fue poblacional. Los tamaños proyectados y logrados aparecen en la tabla 1.2. y, como puede apreciarse, no hay variación entre los centros proyectados y los logrados.

Tabla 1.2. Tamaño de la muestra				
Muestra	Tamaño proyectado	Tamaño logrado		
Alumnado	27600	28708		
Centros	887	887		
Profesorado	1500	1341		
Familias	27600	25741		
Directoras y directores	887	874		



Aunque teóricamente todas las comunidades autónomas tendrían que haber participado con el mismo número de centros, alumnos, familias, profesorado y equipos directivos, su aportación final a la muestra es diferente como se observa en la tabla 1.3.

Tabla 1.3. Tamaño de la muestra lograda por comunidades autónomas

Comunidad	Alumnado	Familias	Profesorado	Directoras y directores
Andalucía	1674	1474	74	50
Aragón	1601	1466	75	49
Principado de Asturias	1532	1442	72	50
Islas Baleares	1642	1437	75	49
Canarias	1677	1500	71	46
Cantabria	1563	1448	75	50
Castilla-La Mancha	1537	1409	74	50
Castilla y León	1389	1290	75	49
Cataluña	1570	1416	72	49
Comunidad Valenciana	1637	1511	75	50
Extremadura	1440	1350	74	49
Galicia	1492	1376	75	50
Comunidad de Madrid	1696	1570	75	50
Región de Murcia	1656	1471	75	50
C. Foral de Navarra	1516	1320	74	49
País Vasco	1518	1114	74	50
La Rioja	1671	1592	75	50
Ceuta	959	819	43	20
Melilla	938	736	38	14

2. FINALIDAD Y OBJETIVOS

El objetivo general de este proyecto es contribuir a la mejora de la calidad de la educación, aumentar la transparencia y la eficacia del sistema educativo y valorar en qué medida la escuela forma a los alumnos para asumir su papel como ciudadanos en una sociedad moderna y dinámica.

Es importante mencionar que no entra dentro de los objetivos de esta evaluación dar información individualizada ni a estudiantes ni a centros.

Los objetivos específicos son averiguar si existen o no otro tipo de relaciones entre las variables, menos intuitivas,



((((Debido al gran número de variables, 450, en lugar de escoger las variables que no queríamos, se han elegido las variables que se consideran relevantes para el proyecto, de acuerdo con los objetivos específicos, descartando el resto para este estudio. Cuando comentemos la limpieza, hacer un anexo explicando que hay 5 códigos para los datos faltantes. Hemos eliminado unos 4551 casos con datos ausentes)))

CALIDAD Y LIMPIEZA DE DATOS

La base de datos de la se disponía desde el inicio estaba formada por 450 variables y unos 28700 casos aproximadamente. Muchos de las observaciones de cada individuo estaban ausentes. Por tanto, se decidió reducir la base de datos eliminando estas observaciones. Debido a que la muestra constaba de 28. 708 alumnos, se ha llegado a la conclusión de que este tipo de limpieza es el adecuado y, aunque se eliminen unos casos, sigue siendo representante de la población real.

Además, se han eliminado variables que no eran útiles para los objetivos propuestos o se entendían, como por ejemplo las variables "Plan de formación – aspectos curriculares" y "Peso", respectivamente.

Es importante destacar que se ha visto la obligación de tener que elegir entre el cuestionario hecho a los padres y el hecho a los hijos, ya que las preguntas a las que debían contestar eran las mismas y discrepancias entre padres e hijos podría causar errores en los análisis. Por tanto, se decidió elegir la encuesta hecha a los padres ya que se consideró que sus respuestas pueden ser las más cercanas a la realidad, ya que las respuestas de los niños podrían no ser del todo fiables debido a su corta edad.

Como consecuencia, la base de datos se ha reducido en 11206 casos y en 384 variables, quedándose un libro de datos de 17494 casos y 66 variables.

A continuación, se realizó la recodificación de los datos y la creación de nuevas variables.



En cuanto a la recodificación,se realizó un cambio de las observaciones cualitativas al valor que representan, es decir, se cambiaron, por ejemplo, los valores "0" de la variable Vive con la Madre por "NO". Así, al realizar un histograma o cualquier otra tabla o análisis, su comprensión será más rápida, ya que sin este cambio se perdería mucho tiempo tratando de hallar lo que representan los valores de la base de datos.

Por último, se crearon nuevas variables que facilitarán nuestro análisis. En este caso, se crearon las variables:

- **Matemáticas:** es la media de las 4 pruebas de matemáticas de las que disponíamos desde un inicio.
- **Lingüística**: es la media de las 4 pruebas de lingüística de las que disponíamos desde un inicio.
- **Mundo Físico:** es la media de las 4 pruebas de mundo físico de las que disponíamos desde un inicio.
- **Lingüística**: es la media de las 4 pruebas de lingüística de las que disponíamos desde un inicio.
- **Social y Ciudadanía:** es la media de las 4 pruebas de social y ciudadanía de las que disponíamos desde un inicio.
- **Nota media ciencias:** creada a partir de las notas de Matemáticas y Mundo Físico.
- Nota media letras: creada a partir de las notas de Lingüística y Social y Ciudadanía.
- Nota media: creada a partir de todas las notas.
- **Lenguas de España en Casa:** se trata de los alumnos que hablan valenciano, catalán, gallego o vasco en sus casas.
- Lenguas extranjeras en Casa: se trata de los alumnos que inglés, árabe, rumano... en sus casas.
- Lenguas de España en Cole: se trata de los alumnos que hablan valenciano, catalán, gallego o vasco en sus colegios.
- Lenguas extranjeras en Cole: se trata de los alumnos que inglés, árabe, rumano... en sus colegios.



ANÁLISIS UNIVARIANTE

<u>Titular</u>

La primera de las variables cualitativas que se estudiará para el análisis univariante es "Titular" la cual puede recibir los valores de colegio público o privado.

"Titular" es una variable cualitativa nominal binaria ya que sólo puede tomar dos valores posibles, por lo cual, se se emplearán frecuencias y diagramas de sectores para realizar el estudio.

Esta variable presentaba en la base de datos inicial valores posibles de 1 y 2, que representaban si era público y privado, respectivamente. Se realizó la recodificación gracias a la orden "buscar y reemplazar" que tiene el programa Excel.

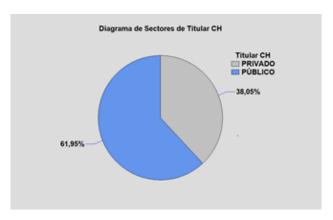


Tabla de Frecuencia para Titular CH

			Frecuencia	Frecuencia	Frecuencia
Clase	Valor	Frecuencia	Relativa	Acumulada	Rel. acum.
1	PRIVADO	6657	0,3805	6657	0,3805
2	PÚBLICO	10837	0,6195	17494	1,0000

En la muestra se observa que el 61,95% asiste a colegios públicos y el resto a colegios privados.

Lengua

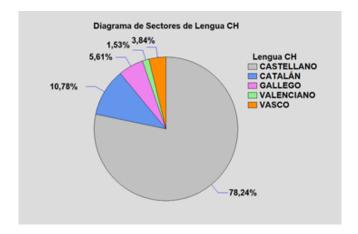
Otra variable importante para el análisis univariante y multivariante es la lengua que hablan los individuos de la muestra. Dicha variable también es cualitativa nominal pero en este caso no es binaria ya que toma 5 valores los cuales son: castellano, catalán, vasco, gallego y valenciano.



Esta variable presentaba en la base de datos inicial valores posibles de 1, 2, 3, 4, 5 y 9, que representaban si esa persona hablaba castellano, catalán, gallego, valenciano, vasco o no ha contestado, respectivamente. Se realizó la recodificación gracias a la orden "buscar y reemplazar" que tiene el programa Excel.

Tabla de Frecuencia para Lengua CH

			Frecuencia	Frecuencia	Frecuencia
Clase	Valor	Frecuencia	Relativa	Acumulada	Rel. acum.
1	CASTELLANO	13687	0,7824	13687	0,7824
2	CATALÁN	1886	0,1078	15573	0,8902
3	GALLEGO	982	0,0561	16555	0,9463
4	VALENCIANO	267	0,0153	16822	0,9616
5	VASCO	672	0,0384	17494	1,0000



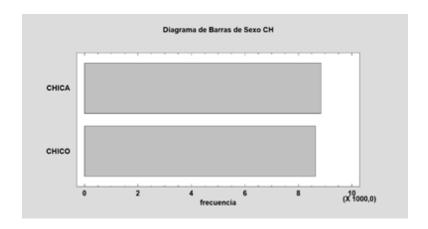
La mayoría de los individuos de nuestra muestra hablan castellano (78,24%) y el 20% restante aproximadamente hablan las otras cuatro lenguas mencionadas anteriormente.

Sexo

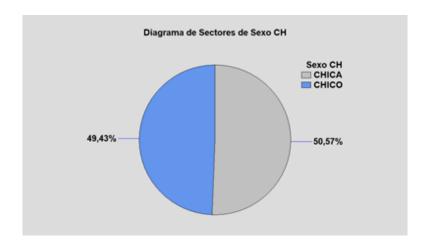
La tercera variable que se estudia es la variable de "Sexo". Dicha variable es nominal binaria ya que simplemente puede tomar los valores de Chico o Chica.

Esta variable presentaba enla base de datos inicial valores posibles de 1 y 2, que representaban si era chica o chico, respectivamente. Se realizó la recodificación gracias a la orden "buscar y reemplazar" que tiene el programa Excel.





En nuestra muestra hay aproximadamente el mismo número de chicos que de chicas.



Repetición de curso

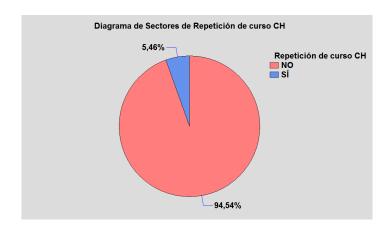
La variable "Repetición de curso" indica si el alumno en cuestión ha repetido curso, tomando dos valores posibles, "sí" o "no".

Esta variable es *cualitativa nominal binaria*, pues toma solo dos valores posibles y, por tanto, se emplearán medidas de resumen como las frecuencias y a gráficos como el diagrama de barras, sectores, etc.

			Frecuencia	Frecuencia	Frecuencia
Clase	Valor	Frecuencia	Relativa	Acumulada	Rel. acum.
1	ИО	16539	0,9454	16539	0,9454
2	SÍ	955	0,0546	17494	1,0000



En la tabla de frecuencia anterior se puede observar que de los 17494 alumnos, 16539 no han repetido nunca curso, mientras que 955 sí lo han hecho.



Tiempo de estudio después de clase

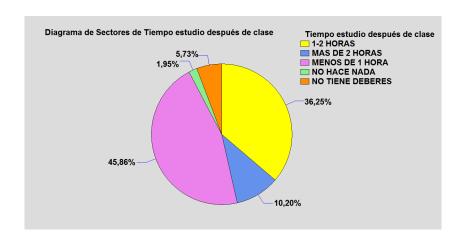
La variable "Tiempo de estudio después de clase" indica el tiempo dedicado al estudio una vez finalizadas las clases del día, esta variable puede tomar los siguientes valores, "1-2 horas", "Más de dos horas", "menos de una hora", "No hace nada", "No tiene deberes".

Esta variable es *cualitativa nominal*, pues toma solo dos valores posibles y, por tanto, se utilizarán medidas de resumen como las frecuencias y a gráficos como el diagrama de barras, sectores, etc.

En la tabla de frecuencias se puede observar que la gran mayoría de los alumnos(6342) le dedican una media de 1-2 horas, mientras que solo 341 no hacen nada, dato que resulta curioso teniendo en cuenta que son alumnos de cuarto de primaria

			Frecuencia	Frecuencia	Frecuencia
Clase	Valor	Frecuencia	Relativa	Acumulada	Rel. acum.
1	1-2 HORAS	6342	0,3625	6342	0,3625
2	MAS DE 2 HORAS	1785	0,1020	8127	0,4646
3	MENOS DE 1 HORA	8023	0,4586	16150	0,9232
4	NO HACE NADA	341	0,0195	16491	0,9427
5	NO TIENE DEBERES	1003	0,0573	17494	1,0000





¿Te gusta ir al colegio?

La variable "¿te gusta ir al colegio?", cuenta con 17494 casos y con 4 posibles valores distintos. Estos son "Nada", "Poco", "Bastante" y "Mucho".

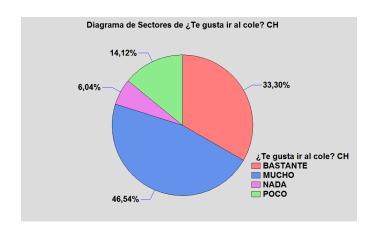
Esta variable es *cualitativa nominal*, pues toma solo dos valores posibles y, por tanto, se emplearán medidas de resumen como las frecuencias y a gráficos como el diagrama de barras, sectores, etc.

La variable anteriormente mencionada proviene de un cuestionario que se le realizó al alumnado.

			Frecuencia	Frecuencia	Frecuencia
Clase	Valor	Frecuencia	Relativa	Acumulada	Rel. acum.
1	BASTANTE	5825	0,3330	5825	0,3330
2	MUCHO	8142	0,4654	13967	0,7984
3	NADA	1057	0,0604	15024	0,8588
4	POCO	2470	0,1412	17494	1,0000

La mayoría de los niños encuestados (8142) han respondido que les gusta "mucho" el colegio, seguido de "bastante" (5825). Estos datos son curiosos ya que el 79.83% de los niños están contentos con ir al colegio, y en la infancia no se suele pensar así.





ISEC (Índice socioeconómico)

Se trata de una variable cuantitativa que indica el nivel económico de los alumnos de la base de datos. Para introducirla, cabe señalar que entre ellos tienen media 0. Así, a medida que se alejen del 0 en el eje x positivo, los alumnos tendrán un mejor economía familiar, y a medida que se alejen de este en el sentido contrario, tendrán una peor.

En este caso, los datos fluctúan entre -3,35 (mínimo) y 1,599 (máximo). La media de esta variable es muy cercana a 0, en este caso es de 0,15, por la razón expuesta con anterioridad.

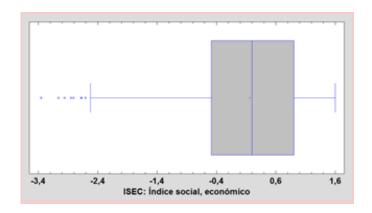
Presenta una asimetría negativa y con una distribución con datos mayoritariamente más "aplanados" de lo normal, como se observa en la curtosis (-5,95) y en el sesgo estandarizado (-24,78).

Todos estos datos están resumidos en el siguiente resumen estadístico:

Recuento	17494
Promedio	0,159621
Desviación Estándar	0,945489
Coeficiente de Variación	592,333%
Minimo	-3,35322
Máximo	1,59986
Rango	4,95308
Sesgo Estandarizado	-24,7826
Curtosis Estandarizada	-5,95201



Además, en el siguiente diagrama de Caja y Bigotes, hallamos que existen datos extremos. Lo que transmite este gráfico, es que existen más alumnos con un ISEC mayor que 0, aunque encontramos los datos extremos a la izquierda de la media. Por tanto, existen alumnos con una economía y un índice social muy pobre comparada con la mayoría de ellos.



Lingüística

Se trata de una variable cuantitativa que nos indica la nota en Lingüística de todos los alumnos. Esta variable, ha sido creada gracias a la media entre las cuatro pruebas que la componían.

En este caso, los datos van de 193,81, como mínimo, hasta 822,923, como máximo. La nota promedio de los niños es de 519,62.

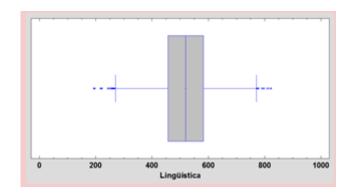
Presenta una distribución normal, con datos mayoritariamente más "aplanados" de lo normal, como podemos observar en la curtosis (-7,915) y en el sesgo estandarizado (-0,12).

Estos datos se recogen en el siguiente resumen estadístico:



Recuento	17494
Promedio	519,62
Desviación Estándar	88,733
Coeficiente de Variación	17,0765%
Minimo	193,81
Máximo	822,923
Rango	629,113
Sesgo Estandarizado	-0,120414
Curtosis Estandarizada	-7,91481

En el diagrama de Box & Whisker que presenta esta variable, se puede ver que existen valores extremos pero no anómalos, ya que son notas que han obtenido los alumnos en esta asignatura. Esto evidencia que existen niños que se le dan mejor los idiomas que a otros y a otros que se le dan peor que a la mayoría. Más del 50% del alumnado se encuentra entre 450 y 600 puntos, ya que las dos cajas centrales están entre este intervalo.



Matemáticas

Se trata de una variable cuantitativa que nos indica la nota en Matemáticas de todos los alumnos. Esta variable, ha sido creada gracias a la media entre las cuatro pruebas que la componían.

En este caso, los datos van de 191,497, como mínimo, hasta 838,284, como máximo. La nota promedio de los niños es de 517,552.

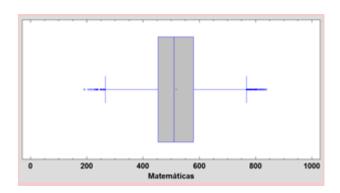
Presenta una asimetría positiva y con una distribución con datos mayoritariamente más "aplanados" de lo normal, como podemos observar en la curtosis (-4,76808) y en el sesgo estandarizado (14,8097).

Todos estos datos se pueden observar en el siguiente resumen estadístico:



Recuento	17494
Promedio	517,552
Desviación Estándar	90,0107
Coeficiente de Variación	17,3916%
Minimo	191,497
Máximo	838,284
Rango	646,787
Sesgo Estandarizado	14,8097
Curtosis Estandarizada	-4,76808

Al igual que en la variable anterior, existen valores extremos pero no anómalos que representan diversas puntuaciones en esta asignatura.



Mundo físico

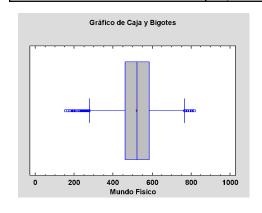
Es una media de distintas áreas que son: la salud, el medio ambiente y las relaciones entre ciencia y tecnología. La nota media es de 520 puntos, observamos que existe una desviación típica de 70 puntos.

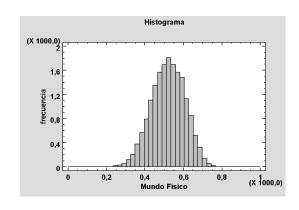
En cuanto a el rango de la variable observamos que va desde 156 puntos a 815. También, podemos notificar que la variable sigue una distribución muy parecida a la normal, centrándose esta alrededor de los 520 puntos.

Por último, existen ciertos valores atípicos tanto en la cola trasera(notas muy bajas), tanto en la cola delantera(notas muy altas).

Resumen estadístico de mundo físico

Resultien estadistico de mundo físico		
Promedio	520,528	
Desviación Estándar	86,8315	
Mínimo	156,742	
Máximo	815.205	







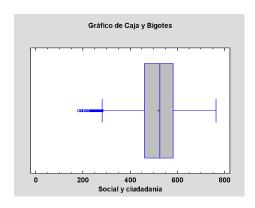
Nota social y ciudadanía

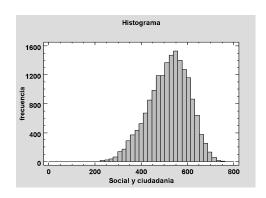
Mide el grado de comprensión de la realidad social en la que viven, de convivencia y de ejecución de la ciudadanía democrática en una sociedad plural y de compromiso a contribuir a su mejora.

Se trata de una variable cuantitativa. Tiene una media de 518 puntos y una desviación típica de 87 puntos. Las notas de social y ciudadanía toman valores comprendidos desde 180 puntos(nota más baja)) a 763(nota más alta).

Observamos que existen ciertos valores atípicos correspondientes a las notas más bajas. Por último podemos observar que sigue una distribución parecida a la normal.

Promedio	518,748
Desviación Estándar	87,5894
Mínimo	180,959
Máximo	763,481





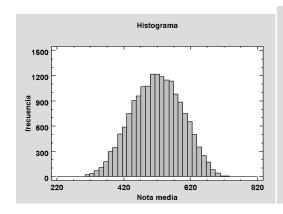
Nota media

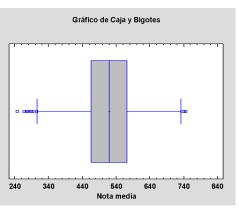
Se trata de una media de las demás notas. Tiene una media de 519 puntos y un desviación típica de 74 puntos. Toma valores comprendidos entre 246 puntos y 746 puntos.

Esta variable sigue una distribución muy parecida a la normal. Por último, observamos que existen ciertos valores atípicos correspondientes a notas muy bajas y muy altas.

Resumen Estadístico para Nota media

Promedio	519,112
Desviación Estándar	74,6452
Mínimo	246,812
Máximo	746,434







VARIABLES PARA ANÁLISIS BIVARIANTE Y MULTIVARIANTE

- Centro(público o privado) con notas
- ISEC con rendimiento niños(notas)
- ¿Te gusta ir al colegio? con notas
- Relación profesores con notas
- Sexo con notas
- variable lengua casa con nota lengua
- ayuda en casa con notas
- tiempo estudio con notas
- numero de lenguas que hablan en casa con notas
- relacion compañeros con notas

Analisis de probabilidades (probabilidad de que un niño...)



RELACIONES

Media total con intervalos está relacionado con Frec Internet CH

Media total con intervalos está relacionado con Casa: Tranquilidad para estudio CH

LENGUAS TOTALES QUE HABLAN está relacionado con Media total con intervalos

Padres van a tutorías CH está relacionado con Media total con intervalos

Padres van a reuniones CH está relacionado con Media total con intervalos

Dificultad enseñanza: material ausente/malo CH está relacionado con Media total con intervalos

Dificultad enseñanza: Falta personal apoyo docente CH está relacionado con Media total con intervalos

Dificultad enseñanza: Falta prof cualif CH está relacionado con Media total con intervalos

Caract Ciudad del cole ${\tt N}^{\circ}$ HAB CH está relacionado con Media total con intervalos

Trabajo PAPÁ CH está relacionado con Media total con intervalos

Estudios MAMÁ CH está relacionado con Media total con intervalos

Tabajo MAMÁ CH está relacionado con Media total con intervalos

Ayuda estudios: PROF PARTICULAR está relacionado con Media total con intervalos

Ayuda estudios: ACADEMIA no está relacionado con Media total con intervalos

Ayuda estudios: PADRES está relacionado con Media total con intervalos

Ayuda estudios: HERMAN@S está relacionado con Media total con intervalos

Repetición de curso CH está relacionado con Media total con intervalos

Tiempo estudio después de clases CH está relacionado con Media total con intervalos

Relación con compañer@s CH está relacionado con Media total con intervalos



¿Te gusta ir al cole? CH está relacionado con Media total con intervalos

Cambio de colegio CH está relacionado con Media total con intervalos

Buena relación: MADRE CH está relacionado con Media total con intervalos

Buena relación: PADRE CH está relacionado con Media total con intervalos

¿Hasta dónde quieres estudiar? CH está relacionado con Media total con intervalos

Frec Libros de lectura CH está relacionado con Media total con intervalos

Frec Enciclopedias CH está relacionado con Media total con intervalos Frec Internet CH está relacionado con Media total con intervalos

Elemento	Aspecto a valorar	Competencia	Peso
	Justifica razonadamente la necesidad del proyecto	CT-05 Diseño y proyecto	
	Establece unos objetivos claros del proyecto	CT-05 Diseño y proyecto	
	Desarrolla las acciones para la consecución de los objetivos	CT-05 Diseño y proyecto	
	Analiza el desarrollo de las actividades a partir de las sugerencias marcadas	CT-12 Planificación y gestión del tiempo	
Memoria	Seleccionar y utilizar las técnicas y herramientas de visualización más adecuadas para representar datos y resultados en un contexto determinado.	C. Específica	50%
	Diseñar e implementar los procesos de selección de datos y de verificación de su calidad a tendiendo fundamentalmente a su heterogeneidad y a su variabilidad.	C. Específica	

