

Ojito con la Estadística



PARTICIPANTES

Virginia Alzás Gómez

Esperanza Becerra Gómez

Lucía Blanco Mata

Esther Fernández Gómez

Marta Gago Mangas



CURSO

1º Bachillerato Científico - Tecnológico y Ciencias de la Salud.

TUTORA

M^a Mar Mota Medina. Profesora de Matemáticas.

AGRADECIMIENTOS

Raquel Muñoz Vara. Profesora de Matemáticas.

CENTRO

IES Virgen de Soterraño (Barcarrota)

1. INTRODUCCIÓN
2. OBJETIVOS
3. METODOLOGÍA
4. PLANTEAMIENTO DEL TRABAJO
 - 4.1. DEFINICIÓN DE CORRELACIONES ESPURIAS.
 - 4.2. EJEMPLOS DE CORRELACIONES ESPURIAS Y EXPLICACIÓN.
 - 4.3. ESTUDIO DE EJEMPLOS PROPIOS.
5. CONCLUSIONES FINALES
6. LIMITACIONES
7. BIBLIOGRAFÍA

1. Introducción

Últimamente es frecuente ver titulares, particularmente sensacionalistas del tipo “científicos encuentran una conexión entre x e y” como:

¿Comer margarina lleva al divorcio?

James Fletcher
BBC

1 junio 2014

f t v e Compartir

"El consumo de margarina está vinculado al divorcio". Si usted viera este titular en un diario o sitio de internet, ¿qué pensaría?

Y ¿qué pasaría si leyera un poco más y encontrara un gráfico convincente que le mostrara que los índices de divorcio y de consumo de margarina coinciden a lo largo de unos 10 años?

Indice de divorcio en Maine - Consumo de margarina per cápita

Correlación: 99%

Indice de divorcio en Maine por 1.000 personas

Consumo de margarina per cápita (liba)

2000 01 02 03 04 05 06 07 08 09 10

Fuente: Census de EE.LU., USDA, Myriagen.com

BBC WORLD SERVICE

“Un estudio afirma que las personas que fuman ligan más”

Un grupo de experimentadores ha comparado el grupo 1 “fumadores” con el grupo 2 “no fumadores” y ha constatado que el grupo 1 tenía un historial de experiencias amorosas mayor que el grupo 2.

Ver la televisión acorta la vida hasta en cinco años

Si además este titular, va acompañado de un gráfico convincente, todos nosotros nos inclinamos a pensar que podría haber un vínculo entre estos hechos.

Pero el problema es que intentamos encontrar una relación **causa - efecto** y podemos concluir que efectivamente, “consumir margarina conduce al divorcio”, “que fumar implica ligan más” o “ver la televisión es un factor de riesgo para nuestra salud”.

Por lo tanto, en esta era digital que proliferan estudios y noticias, debemos ser muy críticos y no dejarnos manipular tan fácilmente. De ahí, que sea esencial tener conocimientos básicos de Estadística. Por ello, nos hemos propuesto profundizar en este tema.

Además, hay que añadir la originalidad que ofrece esta temática, ya que nos posibilita relacionar elementos que, a simple vista, no tienen relación alguna, pero que a la vez, su conexión puede llegar a ser curiosa, o incluso divertida.

También, como consecuencia de lo anterior, es importante e interesante el concepto de *la tercera variable*. Los elementos en cuestión están vinculados aparentemente por dos componentes; sin embargo, detrás de estos últimos, pueden existir diferentes causas que también influyen en la relación.

Por otra parte, nos permite conocer de cerca numerosos aspectos que pasan desapercibidos en nuestra vida cotidiana, favoreciendo así el aumento de nuestra cultura. Por tanto, nos incita a interesarnos por temas que anteriormente habríamos visto de pasada.

¡Pues bien, empecemos!

2. Objetivos

1. Profundizar en el concepto de correlación espuria y buscar ejemplos en medios de comunicación y páginas web.
2. Discernir entre dos conceptos importantes como son la correlación y la causalidad.
3. Determinar mecanismos para evitar el engaño en trabajos estadísticos.
4. Aprender conceptos básicos de Estadística y su aplicación práctica en un proyecto de investigación.

3. Metodología

Hemos extraído información de diversas webs para comentar ejemplos de correlaciones espurias, pero para elaborar nuestros propios ejemplos, hemos acudido al Instituto Nacional de Estadística (INE), a la SEO (Sociedad Española de Ornitología), y de la AEMET (Agencia Estatal de Meteorología) y a fuentes propias como los datos proporcionados por la secretaría de nuestro centro educativo y un cuestionario realizado entre los alumnos del instituto.

Para realizar el análisis estadístico de los datos hemos utilizado la hoja de cálculo de Libre Office. Hemos empleado el filtro de datos para poder hacer el recuento más cómodamente y funciones como PROMEDIO, VARIANZA, DESVIACIÓN, COEF. DE CORRELACIÓN, etc. Y por supuesto, hemos realizado los gráficos con ellas. Y además, se ha utilizado el programa R, para iniciarnos en el aprendizaje de un programa estadístico y así, nos ha permitido hacer el análisis de los datos más rápido.

Por último, la presentación ha sido realizada en Google Drive para facilitarnos el trabajo en equipo.

4. Planteamiento del trabajo

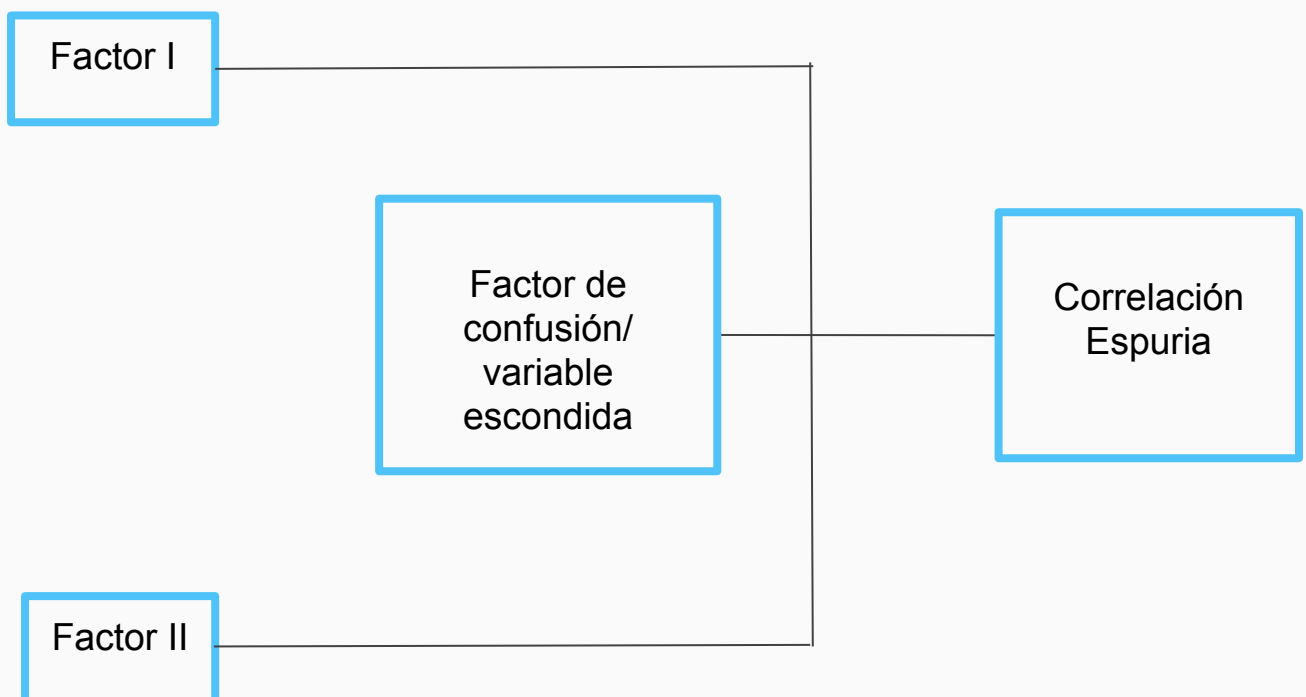
Nos hemos propuesto las siguientes fases en este trabajo:

1. DEFINICIÓN DE CORRELACIONES ESPURIAS.
2. EJEMPLOS DE CORRELACIONES ESPURIAS Y EXPLICACIÓN.
3. ESTUDIO DE EJEMPLOS PROPIOS.

4.1 Definición de correlación espuria

En estadística, se le denomina **relación o correlación espuria** a una relación matemática, en la que se relacionan dos acontecimientos, los cuales no tienen conexión lógica, aunque se puede implicar que la tienen debido a un tercer factor no considerado de forma inmediata, llamado “factor de confusión” o “variable escondida”.

La relación espuria, por tanto, ofrece la impresión de existencia de un vínculo apreciable entre dos grupos, el cual es inválido cuando se examina objetivamente.



4.1 Definición de correlación espuria

Las correlaciones no implican causalidad, es decir, que exista un coeficiente de correlación alto entre dos variables no implica necesariamente que una variable sea causa de la otra.

El hecho de que dos eventos se den habitualmente de manera consecutiva, no implica que uno sea causa del otro. Así, por ejemplo, cuando llueve es más probable que truene, pero no es la lluvia la que causa los truenos.



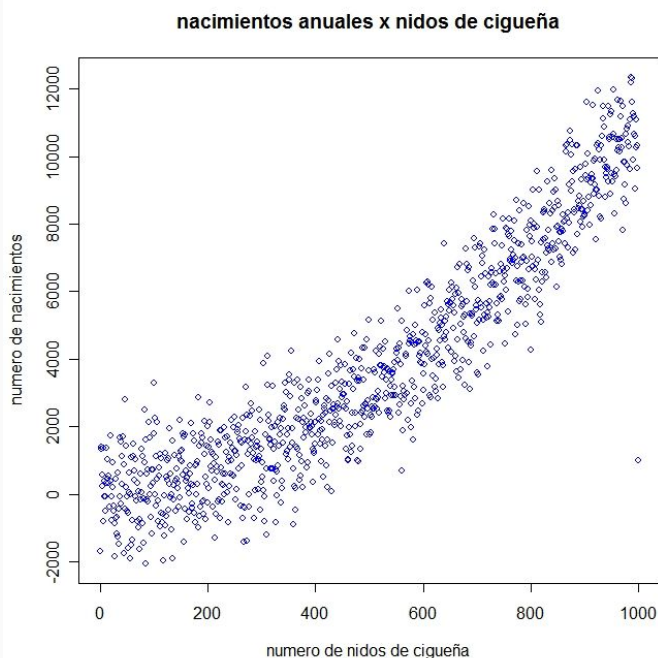
De todos modos, hay numerosas ocasiones que nos llevan a pensar que sí existe una causa responsable de dicha relación. Sin embargo, no la hay, sino que pueden ocurrir diferentes situaciones:

- A causa a B.
- B causa a A.
- La existencia de un tercer fenómeno, C.
- Azar.

4.2 Ejemplos de correlaciones espurias y explicación

A continuación pasamos a describir estudios que han encontrado correlaciones, como mínimo curiosas, entre dos variables y que pueden llevar a confusión, si pensamos que correlación implica causalidad.

Ejemplo 1.- Correlación entre el número de cigüeñas y el número de nacimientos

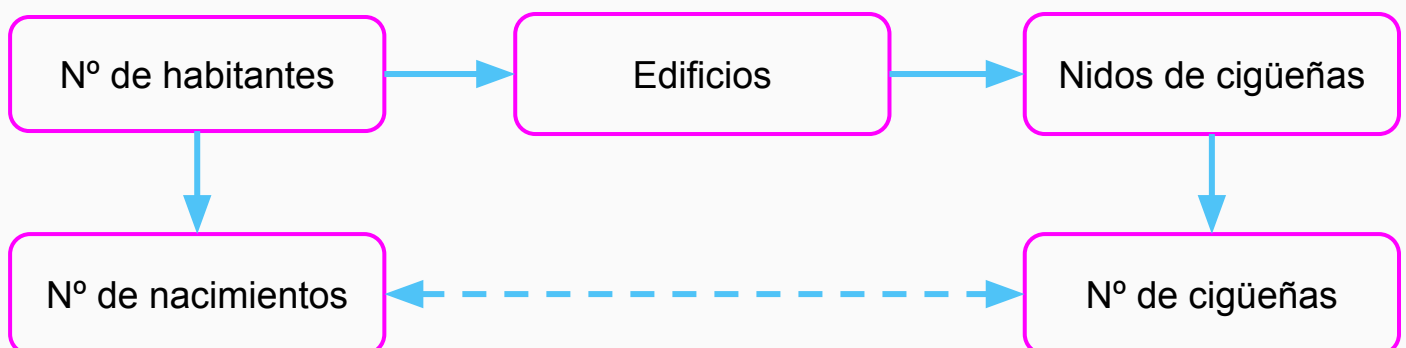


A simple vista puede parecer que no están relacionadas, pero ambas dependen de una tercera variable: **el número de habitantes.**

Cuanto más habitantes hay, la construcción de edificios (iglesias, campanarios, pisos...) es mayor y por lo tanto aumenta la cantidad de cigüeñas.

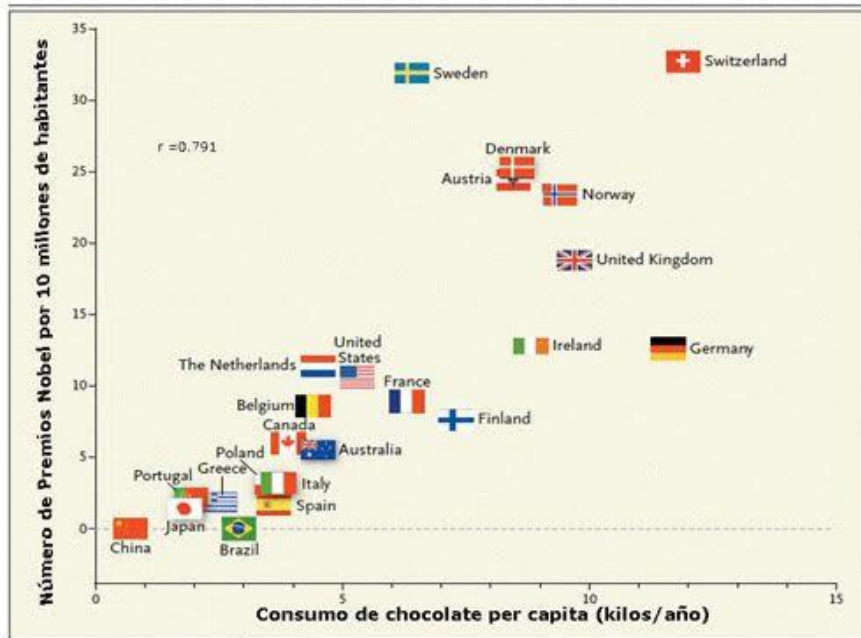
A su vez, al haber más habitantes, también se incrementa el número de nacimientos.

Ahora vamos a explicar esta correlación con un grafo.



4.2 Ejemplos de correlaciones espurias y explicación

Ejemplo 2.- Correlación entre los premios Nobel y el consumo de chocolate

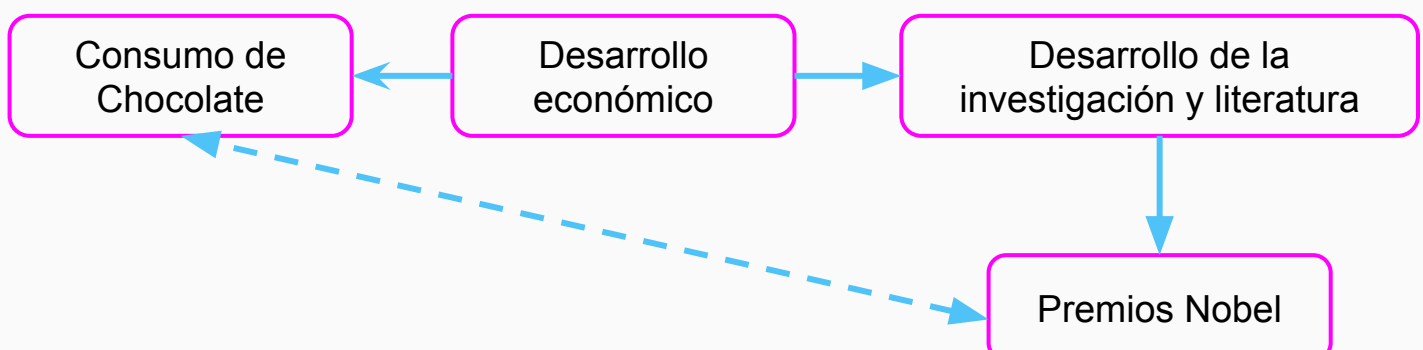


Según la gráfica el número de premios Nobel aumenta si se incrementa el consumo de chocolate. En un primer vistazo puede parecer que esto no tiene ningún sentido, pero su explicación puede radicar en el desarrollo económico.

En los países en los que se da dicho desarrollo, aumenta el consumo de chocolate al igual que aumenta el consumo del resto de alimentos puesto que se dispone de los fondos económicos necesarios para que esto ocurra.

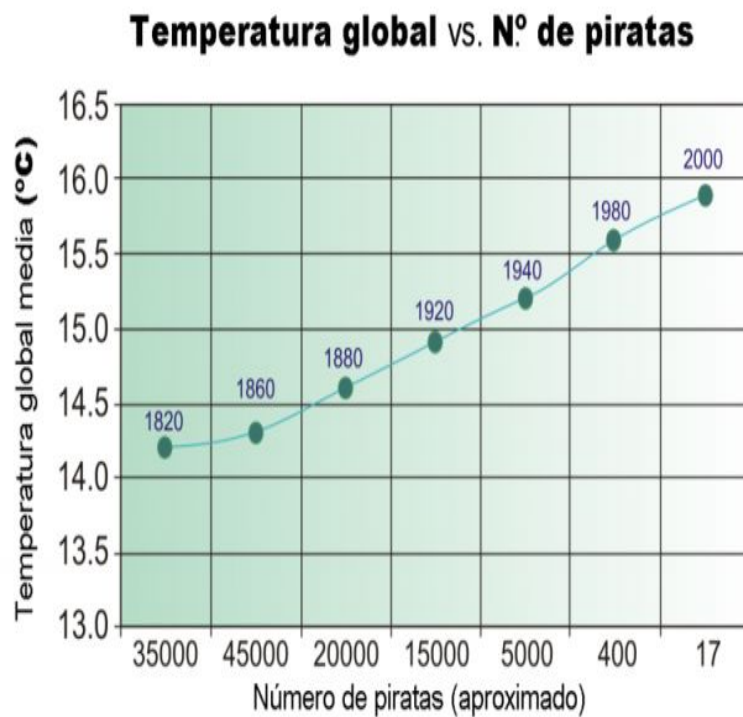
Del mismo modo, una buena situación económica permite el desarrollo de la investigación, literatura... y aumentan las posibilidades de conseguir un premio Nobel.

Ahora vamos a intentar encontrar la explicación a esta correlación con un grafo



4.2 Ejemplos de correlaciones espurias y explicación

Ejemplo 3.- Correlación entre el nº de piratas y la temperatura global

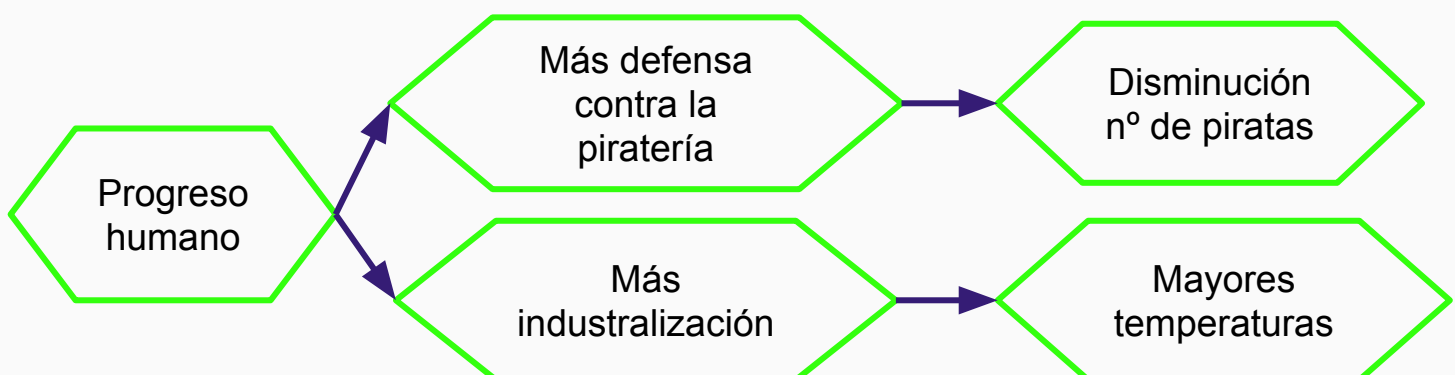


A partir de esta gráfica, podríamos concluir que la disminución del número de piratas ha provocado el cambio climático.

Sin embargo, sería más lógico pensar que actualmente, los Estados pueden combatir mejor la piratería y por eso, seguir surcando los mares no tiene tanto sentido como en el s. XIX, y por otra parte, la revolución industrial y el desarrollo conlleva contaminación, lo que supone el aumento de gases que producen el efecto invernadero.

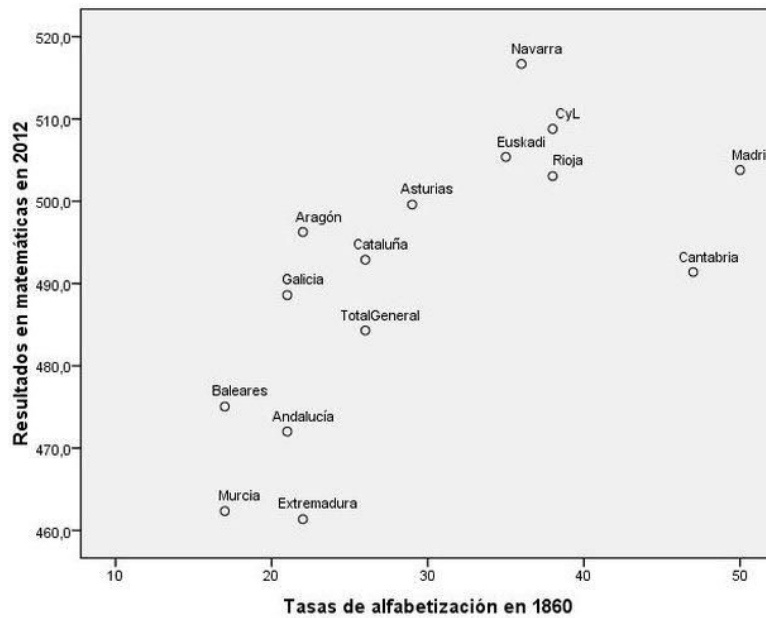
Es decir, que ambas variables parecen ser consecuencias del progreso humano.

A continuación, explicaremos la correlación mediante un grafo.



4.2 Ejemplos de correlaciones espurias y explicación

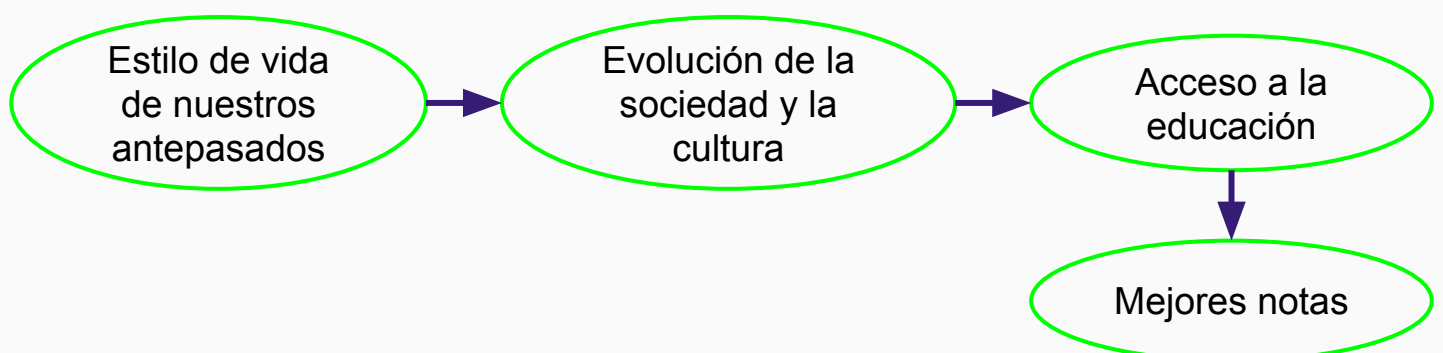
Ejemplo 4.- Correlación entre los resultados en matemáticas en 2.012 y la tasa de alfabetización en 1.860



Ambas variables están relacionadas con el desarrollo de la cultura. El estilo de vida de nuestros antepasados se basaba principalmente en la agricultura y la ganadería, por lo que no tenían tiempo para estudiar. Conforme pasaron los años y la sociedad evolucionó la gente se dedicó a los estudios y por consiguiente se fueron consiguiendo mejores resultados académicos. Concluimos en que quizás esta sea la causa de que actualmente haya una mayor tasa de alfabetización y también mejores resultados en matemáticas.

Además, cabe añadir que en lugares como Extremadura o Murcia, ambas variables son bajas debido a que se tratan de entornos rurales en los que la mayoría de la población se dedica a las actividades anteriormente dichas.

Seguidamente vamos a explicar la correlación mediante un grafo.



4.3 Estudio de ejemplos propios

A continuación vamos a estudiar tres casos de correlaciones espurias, los cuales, intentaremos dar una explicación matemática, buscando una tercera variable que pueda ser la conexión lógica entre ambas variables.

Caso 1.- ¿Traen las cigüeñas a los bebés?

Vamos a estudiar si existe correlación entre el número de cigüeñas y número de nacimientos en una población.

Para ello, hemos obtenido los datos de población del instituto nacional de estadística (INE). En segundo lugar y refiriéndonos a los datos del número cigüeñas, los hemos conseguido en la página web SEO.org.

Cabe añadir que nos hemos centrado en los datos de 2004, pues es el único año del que teníamos información suficiente.



The screenshot shows the INE website with a table of indicators for 2004. The table includes the following data:

Indicador	Periodo	Valor	Variación (%)
IPC	2017M04	—	2,6
EPA. Ocupados (miles)	2017T1	18.438,3	2,27
EPA. Tasa de paro	2017T1	18,75	-2,25
PIB	2017T1	—	3,0
Población total (miles)	2016	46.468,1	0,12



Una vez conseguidos los datos, hemos calculado el coeficiente de correlación de Pearson mediante una tabla de frecuencias y el cálculo de las medidas estadísticas necesarias:

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}$$

4.3 Estudio de ejemplos propios

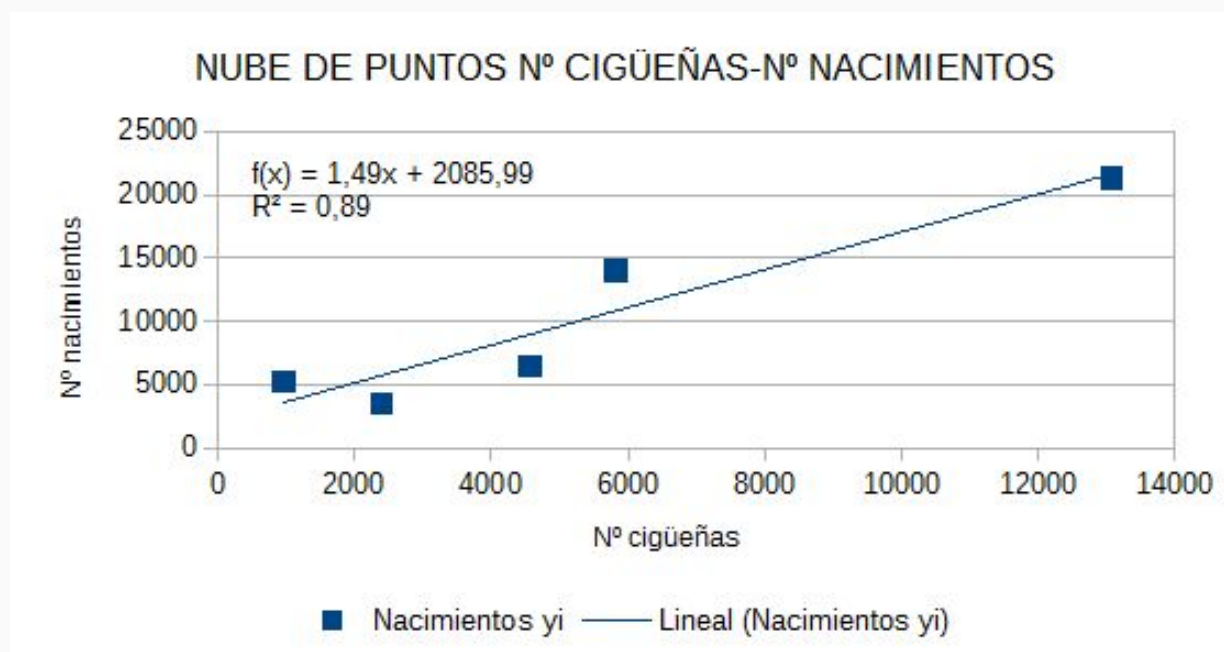
Caso 1.- ¿Traen las cigüeñas a los bebés?

Esta es nuestra tabla de frecuencias realizada con la hoja de cálculo:

	Censo cigüeñas xi	Nacimientos yi	xi ²	yi ²	xi*yi
06 Badajoz	4564	6397	20830096	40921609	29195908
10 Cáceres	2388	3.502	5702544	12264004	8362776
11 Cádiz	5825	14.040	33930625	197121600	81783000
21 Huelva	968	5.237	937024	27426169	5069416
41 Sevilla	13075	21.295	170.955.625	453477025	278432125
Total	26820	50471	232.355.914	731210407	402843225

MEDIA xi	5364	MEDIA yi	10094,2	COEF. DE PEARSON	0,94
VARIANZA xi	17698686,8	VARIANZA y	44349207,76	RECTA DE REGRESION DE Y SOBRE X	
D.T xi	4206,98	D.T yi	6659,52	$y - 10094,2 = (26423356,2 / 17698686,8) * (x - 5364)$	
				$y - 10094,2 = 1,49 * (x - 5364)$	

Tras averiguar el coeficiente de correlación hemos observado que es muy alto (**0,94**); por tanto, podemos concluir que existe una correlación positiva, a medida que aumenta el número de cigüeñas se eleva la cantidad de nacimientos.



En este gráfico podemos observar cómo los puntos se adaptan a la línea. Vemos que los puntos están relativamente alineados; así pues, esta es la comprobación del resultado que hemos obtenido con el coeficiente de correlación.

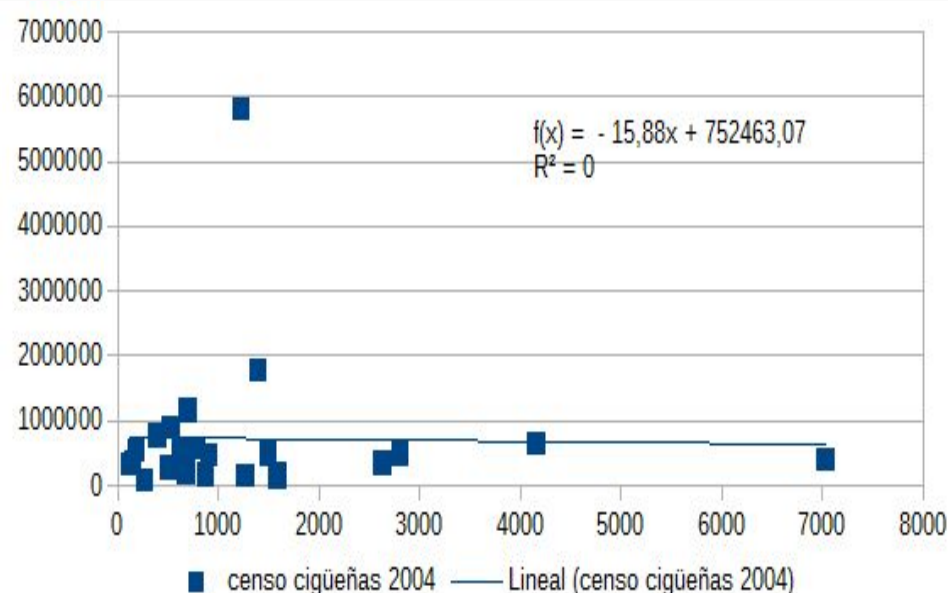
4.3 Estudio de ejemplos propios

Caso 1.- ¿Traen las cigüeñas a los bebés?

Sin embargo, no podemos concluir que la correlación implique causa. Y además, en nuestro caso, la explicación de este altísimo coeficiente radica en que **hemos seleccionado un número muy escaso de datos**, en concreto 5 provincias, lo que facilita que sean posibles estos resultados. Debemos destacar que hemos escogido estas provincias porque son lugares cálidos y es donde más cigüeñas se encuentran en la estación de invierno, periodo en el que se han tomado los datos utilizados.

Si repetimos el estudio con muchos más datos se pierde toda la relación entre ambas variables.

Así lo podemos observar en este último gráfico: en él es imposible apreciar una pequeña alineación de los puntos. Este gráfico es un resultado lógico si tenemos en cuenta que el coeficiente de correlación es de -0,02, es decir, no hay relación.



	Censo cigüeñas	Población
05 Ávila	1261	166.108
06 Badajoz	4155	663.896
10 Cáceres	7035	411.390
11 Cádiz	692	1.164.374
39 Cantabria	177	554.784
13 Ciudad Real	1493	492.914
14 Córdoba	391	779.870
21 Huelva	904	476.707
22 Huesca	670	212.901
24 León	2799	492.720
27 Lugo	142	358.452
28 Madrid	1221	5.804.829
31 Navarra	621	584.734
32 Ourense	112	340.258
34 Palencia	866	173.990
26 Rioja, La	499	293.553
37 Salamanca	2627	350.984
40 Segovia	1582	152.640
41 Sevilla	1391	1.792.420
42 Soria	261	91.652
45 Toledo	777	578.060
47 Valladolid	663	510.863
49 Zamora	1581	198.524

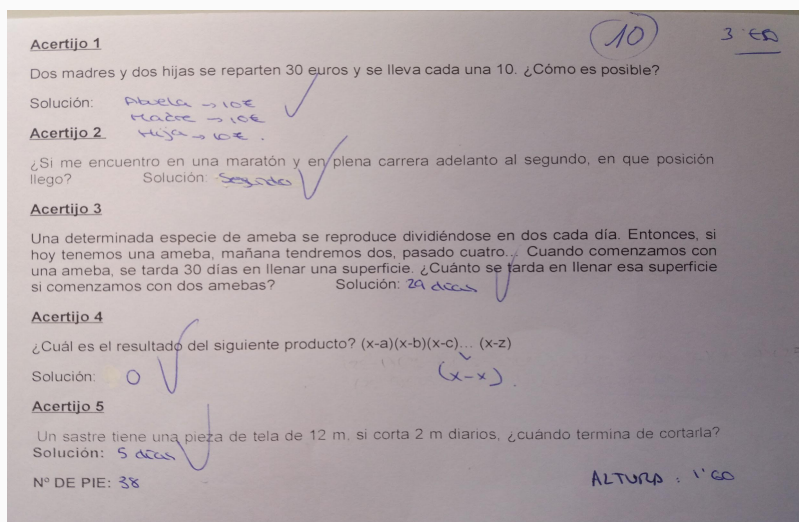
4.3 Estudio de ejemplos propios

Caso 2.- ¿Las personas más altas son más inteligentes?

A continuación vamos a determinar si existe relación entre el ingenio y la altura. Para ello, diseñamos un pequeño test de ingenio, el cual consta de 5 acertijos y además, preguntamos por el nº de pie y su altura.

Lo pasamos a 123 alumnos, eligiendo al azar una clase de cada nivel educativo.

Hemo obteniendo así diferentes resultados. Este es uno de los más sorprendentes:



Obtenidos los datos, calculamos el coeficiente de correlación pero en este caso, para hacerlo, se utilizó las funciones de la hoja de cálculo. (COEF.DE.CORREL)

Curso	Número de pie	Altura	Resultado
1ºESO	39	140	0
1ºESO	37	142	0
1ºESO	37	145	2
1ºESO	36	148	0
2ºESO	37	148	2
1ºESO	37	150	0
1ºESO	39	150	0
2ºESO	38	150	0
1ºESO	36	152	0
1ºESO	38	152	1
1ºESO	35	152	1
3ºESO	37	152	2
2ºESO	38	152	2
1ºESO	40	153	2
1ºESO	34	153	0
2ºESO	37	153	0
2ºESO	37	153	0
2ºESO	36	154	2
1ºESO	38	155	2
1ºESO	36	155	2
2ºESO	36	155	2
2ºESO	38	155	2
2ºESO	38	155	2
2ºESO	38	155	2
2ºESO	40	155	2
1ºESO	37	156	0
1ºESO	39	156	2
4ºESO	36	156	3
2ºESO	40	156	0
3ºESO	37	157	0
2ºESO	36	157	3

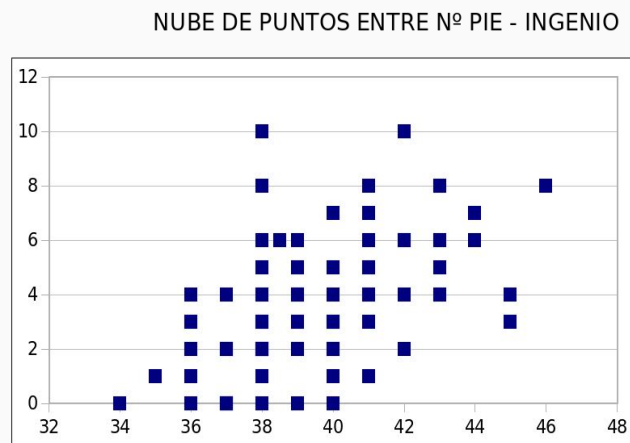
Asistente de funciones
Funciones: Estructura
Categoría: Estadística
Función: COEF.DE.CORREL
Devuelve el coeficiente de correlación entre dos conjuntos de datos.
Datos_1 (obligatorio): C2:C123
Datos_2 (obligatorio): D2:D123
Resultado: 0,8352117581
Fórmula: =COEF.DE.CORREL(C2:C123;D2:D123)

4.3 Estudio de ejemplos propios

Caso 2.- ¿Las personas más altas son más inteligentes?

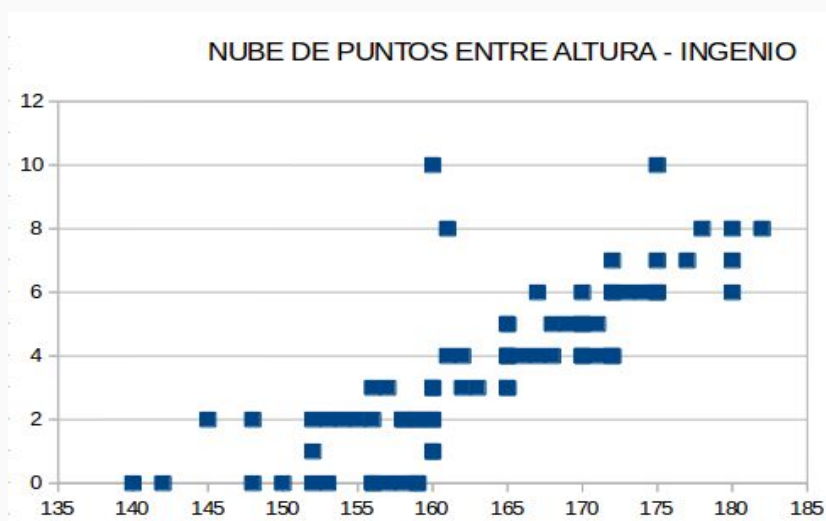
Primero, calculamos el coeficiente de correlación de Pearson entre el **número de pie** y los **resultados en el test**, obteniendo **0,59**, por lo que no podemos afirmar que exista relación entre ambas variables.

Esto también se aprecia si dibujamos la nube de puntos:



A continuación, vamos a estudiar si hay correlación entre **la altura** y los **resultados del test de ingenio**. En este caso, el coeficiente de correlación de Pearson es **0,84**, por lo que podemos afirmar que existe una correlación alta positiva entre las dos variables.

Como nos ha salido un coeficiente de correlación muy alto, vamos a analizar cada curso por separado para intentar buscar una explicación a este hecho, tal vez sea debido a la diferencia de edad.



4.3 Estudio de ejemplos propios

Caso 2.- ¿Las personas más altas son más inteligentes?

- En 1º ESO ocurre lo siguiente:

ESTATURA	0	1	2	3	4	5	6	7	8
<1,5 m	3	0	1	0	0	0	0	0	0
[1,5-1,6)	7	2	7	0	0	0	0	0	0
[1,6-1,7)	0	4	0	6	1	2	0	0	0
[1,7-1,8)	0	0	0	0	0	0	0	0	0
[1,8-1,9)	0	0	0	0	0	0	0	0	0

- En 2º ESO obtenemos los siguientes resultados:

ESTATURA	0	1	2	3	4	5	6	7	8	9	10
<1,5 M	0	0	1	0	0	0	0	0	0	0	0
[1,5 – 1,6)	4	0	8	1	0	0	0	0	0	0	0
[1,6 – 1,7)	0	0	4	2	6	0	0	0	0	0	0
[1,7 – 1,8)	0	0	0	0	1	0	2	0	0	0	0
[1,8 – 1,9)	0	0	0	0	0	0	0	0	1	0	0

- En 3º ESO observamos estos datos:

ESTATURA	0	1	2	3	4	5	6	7	8	9	10
<1,5	0	0	0	0	0	0	0	0	0	0	0
[1,5-1,6)	1	0	2	0	0	0	0	0	0	0	0
[1,6-1,7)	0	0	1	6	2	0	0	0	1	0	1
[1,7-1,8)	0	0	0	0	1	2	1	0	1	0	0
[1,8-1,9)	0	0	0	0	0	0	1	1	0	0	0

4.3 Estudio de ejemplos propios

Caso 2.- ¿Las personas más altas son más inteligentes?

- En 4º ESO recogemos los siguientes resultados:

ESTATURA	0	1	2	3	4	5	6	7	8	9	10
<1,5 m	0	0	0	0	0	0	0	0	0	0	0
[1,5 – 1,6)	0	0	2	1	0	0	0	0	0	0	0
[1,6 – 1,7)	0	0	0	1	3	1	0	0	0	0	0
[1,7 – 1,8)	0	0	0	0	2	1	4	1	0	0	0
[1,8 – 1,9)	0	0	0	0	0	0	0	0	3	0	0

- Y por último, en 2º Bachillerato adquirimos los siguientes datos:

ESTATURA	0	1	2	3	4	5	6	7	8	9	10
<1,5m	0	0	0	0	0	0	0	0	0	0	0
[1,5 – 1,6)	0	0	0	0	0	0	0	0	0	0	0
[1,6 – 1,7)	0	0	2	1	1	3	1	0	0	0	0
[1,7 – 1,8)	0	0	0	0	1	2	4	2	0	0	1
[1,8 – 1,9)	0	0	0	0	0	0	0	0	0	0	0

Por lo tanto, estudiando los datos distinguiendo por curso, observamos que a medida que la edad aumenta, la puntuación del test de ingenio supera el 5, siendo mejores los resultados en las personas de mayor edad.

Luego, la realidad es distinta a lo que podríamos suponer en un principio, ser más alto no causa obtener mejores resultados. Podemos considerar que hay una tercera variable, la **edad**.

A mayor altura, los alumnos suelen tener también más edad, lo que conlleva a tener una mayor madurez y unas capacidades más desarrolladas en los que a estudios se refiere .

4.3 Estudio de ejemplos propios

Caso 3.- ¿El frío activa las neuronas?

¿Cuántos de nosotros creemos que la causa responsable de que en los países situados al norte se obtengan mejores resultados académicos está relacionada con factores meteorológicos?

Es común enlazar los datos académicos de los países del norte al tiempo que en ellos se da; así, al tener un clima más frío, no tienen ganas de salir a la calle, teniendo como distracción el estudio.

A continuación, hemos elaborado un mapa estadístico que refleja el porcentaje de fracaso escolar en Europa y con él, pretendemos reforzar la idea generalizada de que los países más fríos obtienen mejores resultados académicos debido al tiempo.

Los datos en los que nos hemos basado pertenecen a la siguiente página web:



FRACASO ESCOLAR (%) EN LOS PAÍSES DE EUROPA

publicdata
Hacia las sociedades del conocimiento

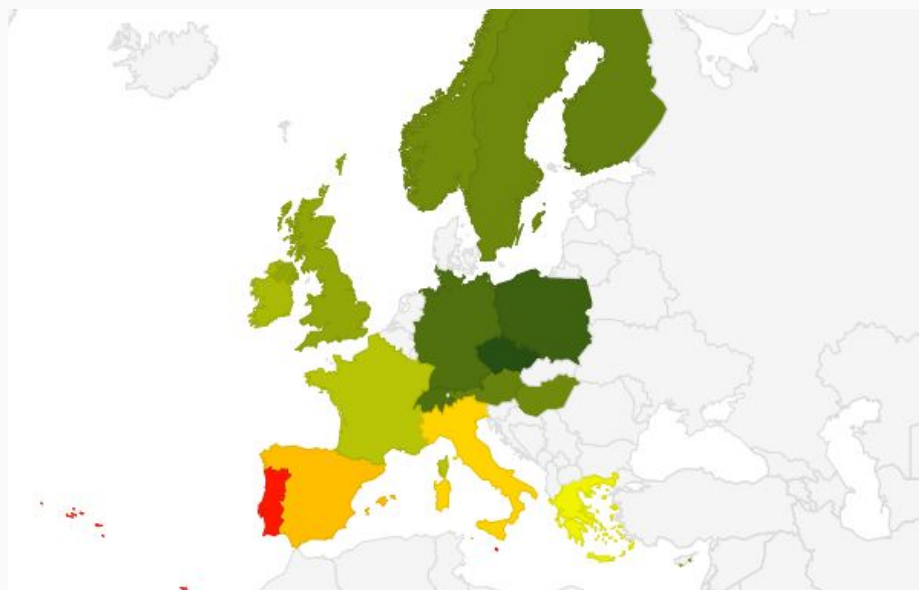
La Comisión Europea considera la educación secundaria superior (equivalente al bachillerato o FP de grado medio en España) como "la formación mínima necesaria para participar en la vida social y económica de un país".
Así pues hemos calculado el llamado "fracaso escolar"* para los países de Europa como el porcentaje de la población adulta (25-64 años) que no ha completado dichos estudios.

* En algunos medios se denomina "fracaso escolar" al [abandono temprano de la educación y formación](#), que es el porcentaje de personas de 18-24 años que no están escolarizadas y que tienen como estudios máximos la educación secundaria o anteriores niveles educativos.

4.3 Estudio de ejemplos propios

Caso 3.- ¿El frío activa las neuronas?

País	% Fracaso escolar
Alemania	14,2
Austria	17,5
España	47,4
Finlandia	17
Francia	29,2
Grecia	37,5
Italia	44,8
Noruega	19,1
Portugal	68,1
República Checa	8,1
Reino Unido	23,9
Suecia	18,4
Suiza	14,2
Irlanda	27,3
Hungría	18,7
Chipre	25,9
Malta	71
Polonia	11,3



Si solo nos muestran este gráfico, interpretaríamos a primera vista que los países del norte, con temperaturas medias más bajas que el resto, tienen menos fracaso escolar que los países europeos restantes.

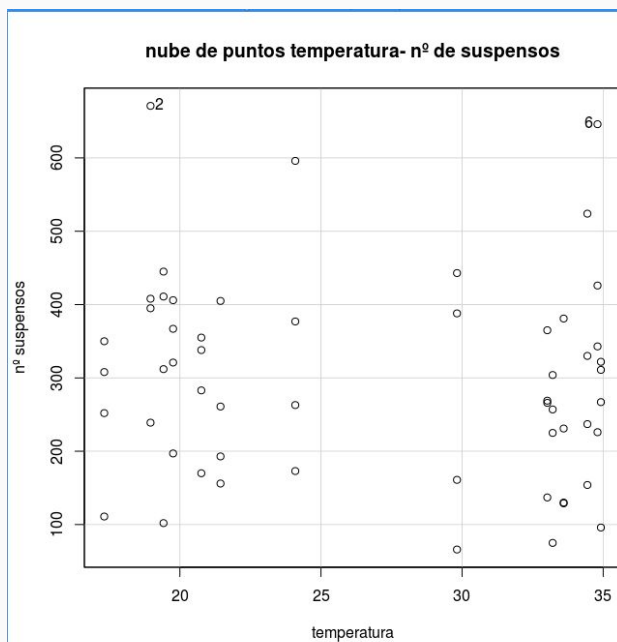
Sería necesario incluir, aparte de estos datos, las temperaturas medias de los países, para poder así realizar una correlación que verifique o no nuestra idea.

4.3 Estudio de ejemplos propios

Caso 3.- ¿El frío activa las neuronas?

Ahora hemos estudiado nuestros datos académicos (desde el año 2008 al 2014) junto con la temperatura media de Barcarrota durante las evaluaciones, para poder comprobar si es verdad que la temperatura influye en nuestro rendimiento escolar.

Para ello utilizaremos el programa de estadística R.



Primero, hemos realizado un gráfico de dispersión de las variables temperatura - nº de suspensos.

Como podemos observar, y se muestra en la nube de puntos, no existe relación alguna entre la temperatura y los resultados académicos.

Calculamos el coeficiente de Pearson, **(-0.12)** el cual apoya el gráfico anterior, sacando como conclusión que no existe relación entre las temperaturas y los datos académicos obtenidos.

```
> cor.test(Datos$Col7, Datos$Suspensos, alternative="two.sided", method="pearson")
```

Pearson's product-moment correlation

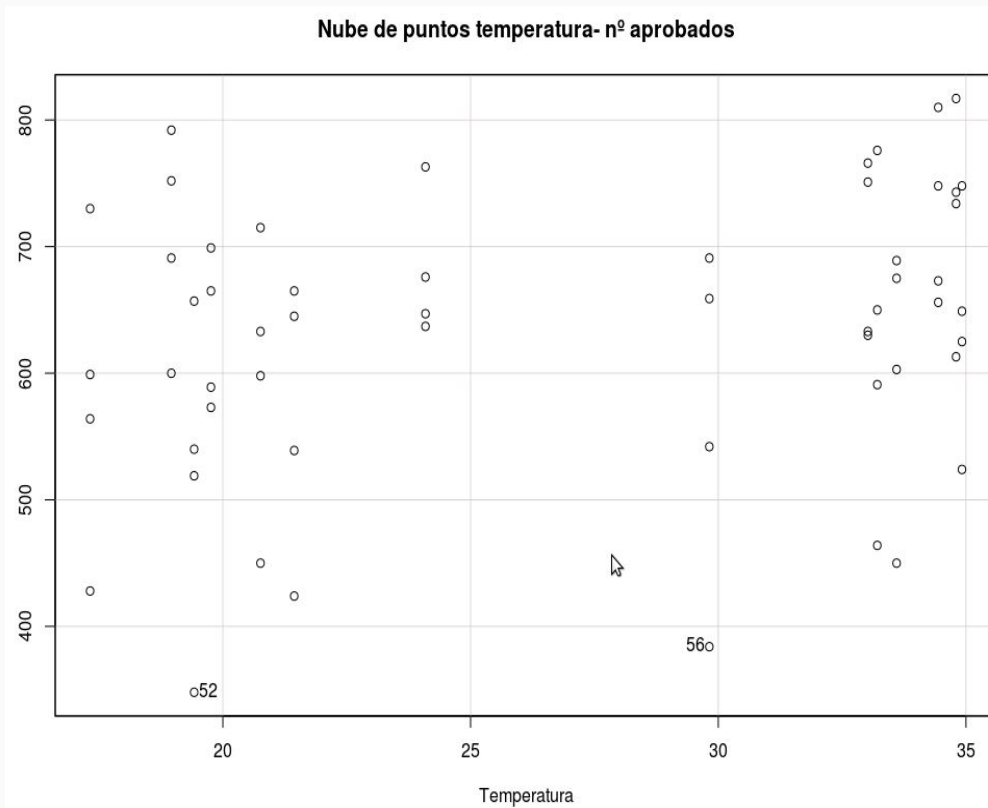
```
data: Datos$Col7 and Datos$Suspensos
t = -0.92134, df = 54, p-value = 0.361
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.3750395 0.1431783
sample estimates:
cor
-0.1244051
```

Por último, pedimos a R el resumen numérico de la variable Nº de suspensos distinguiendo entre las evaluaciones. Existe diferencias en las notas, pero la causa no son las temperaturas, sino que existe una tercera variable: la explicación más lógica es que en la tercera evaluación tendemos a esforzarnos más, como muestra la imagen siguiente:

	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
segunda	316.6071	132.1922	169.00	102	228.50	316.5	397.5	671	28
tercera	275.3214	137.2068	189.25	66	159.25	266.5	348.5	646	28

4.3 Estudio de ejemplos propios

Caso 3.- ¿El frío activa las neuronas?



```
> cor.test(Datos$Aprobados, Datos$Col7, alternative="two.sided", method="pearson")

Pearson's product-moment correlation

data: Datos$Aprobados and Datos$Col7
t = 1.8925, df = 54, p-value = 0.0638
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.01445506  0.48077167
sample estimates:
      cor
0.2493932
```

Esta mínima correlación la podemos comprobar en el coeficiente de Pearson, mostrado en la última imagen. Es un número muy bajo y positivo lo que indica que a mayor temperatura más aprobados, pero la correlación es muy débil. **(0.25)**

Como antes, la diferencia en el rendimiento académico entre la segunda evaluación (que suele realizarse en Marzo) a la tercera evaluación (en Junio), no se debe a la temperatura, sino a que los estudiantes se esfuerzan más en la última evaluación.

5. Conclusiones

La primera conclusión que obtenemos, y de las que derivan todas las demás, es que **correlación no implica causalidad**. Así lo podemos observar teniendo en cuenta varios aspectos:

Por ejemplo, cuando vimos el caso de las cigüeñas, aparentemente parecía que el número de cigüeñas implicaba un mayor número de nacimientos, pero la explicación de esto la podemos encontrar en la **manipulación al elegir los datos**, pues solo hemos seleccionado lugares cálidos donde las cigüeñas pasan el invierno. Es decir, escoger solo **los datos que nos interesan** puede dar lugar a que exista una altísima correlación. Si a esto le sumamos la **escasez de datos**, es más fácil llegar a una correlación lineal ya que es mucho más sencillo alinear un número muy pequeño de puntos que un número muy grande.

Otra conclusión que sacamos es que la presencia de una **variable oculta** puede llevarnos a resultados sorprendentes. En el segundo caso que nosotras hemos estudiado, según nuestros cálculos, los resultados del test de ingenio de cada alumno/a mejoran a medida que aumentan la altura de los mismos/as. Percibimos también un dato importante: a medida que la edad aumenta, los datos académicos superan el 5, siendo mejores resultados en las personas de mayor edad. Por lo tanto, la realidad es bien distinta a lo que hemos afirmado con nuestro trabajo, ya que la altura de una persona no va relacionada con sus resultados del test y viceversa. Podemos considerar así una **tercera variable**, la cual puede ser que a mayor altura, la mayoría de los alumnos/as más altos/as tienen también más edad, lo que conlleva a tener una mayor madurez y unas capacidades más desarrolladas en los que a estudios se refiere. El tema de las variables ocultas puede ser usado en **medios de comunicación** para resaltar ideas que no son las estadísticamente correctas.

5. Conclusiones

Otro aspecto a tener en cuenta son las **ideas prefijadas**, mediante una **mala utilización del método científico**, es decir, enfocar el estudio a los resultados que queremos obtener. Por ejemplo, en el caso de las cigüeñas, elegimos las cinco provincias en las que más tránsito de cigüeñas pensábamos que había, para que así los resultados fueran acordes con lo que queríamos obtener. Otro caso, ha sido en el estudio de la altura y la inteligencia, en el que las preguntas formuladas en el test, las hemos elegido de tal forma que resultaran más complicadas para los niños pequeños y por lo tanto, que no acertaran demasiadas. Así a los mayores les resultaba relativamente más fáciles y podían acertar un mayor número de cuestiones.

También, cabría destacar el uso de las **ideas preconcebidas**, podemos decir que la mayoría de las veces tendemos a pensar, por ejemplo, que en los países del norte sus mejores resultados académicos se deben al tiempo, dado que el frío hace que la gente salga menos de casa y por tanto pasen más tiempo estudiando. Sin embargo, como hemos comprobado, esto no tiene ninguna relación, ya que hay que tener en cuenta otras variables.

Resumiendo, cada vez que nos presenten un estudio o trabajo estadístico, antes de llegar a conclusiones precipitadas, debemos analizar la situación y comprobar si hay un suficiente número de datos, si el estudio se puede extrapolar a otras zonas o otros sujetos, si existe una tercera variable que pueda explicar las conclusiones obtenidas, si los gráficos no se muestran de manera distorsionada para influir en nuestra percepción o si nos dejamos llevar por ideas preconcebidas.

Así que:

¡OJITO CON LA ESTADÍSTICA!

6. Limitaciones

No es fácil encontrar correlaciones espurias. Esto lo comprobamos cuando intentamos buscar una correlación entre el número de pie y el test de ingenio. Nuestra hipótesis era que los niños que estuviesen en cursos inferiores tendrían un pie más pequeño, no obstante estábamos equivocadas. Esto hizo que tuviésemos que cambiar la variable a correlacionar e introducir la altura.

Por otra parte, tampoco es fácil conseguir relaciones sorprendentes sin un trabajo o dedicación previa.

Finalmente, otra limitación ha sido la falta de tiempo con el que hemos podido buscar o pensar posibles relaciones espurias que estuvieran a nuestro alcance para trabajar y sobre las que pudiésemos obtener datos. No es un trabajo de estadística normal porque en ciertos momentos hemos tenido que trabajar “al revés” es decir, buscar los resultados finales para encontrar los resultados y poder estructurar el trabajo como queríamos.

7. Bibliografía

- Bascós i Excruela, E., Pena i Terrén, Z. *Matemáticas aplicadas a Ciencias Sociales. 1º Bachillerato*. Ed. Oxford
- *Estadística para todos* :<http://www.estadisticaparatodos.es/curriculo/descriptiva.html>
- *Naukas*: <http://naukas.com/2012/08/01/correlacion-no-implica-causalidad/>
- *Ciencia explicada*:
<http://www.ciencia-explicada.com/2013/06/correlacion-causalidad-y-grafos-lo-mas.html>