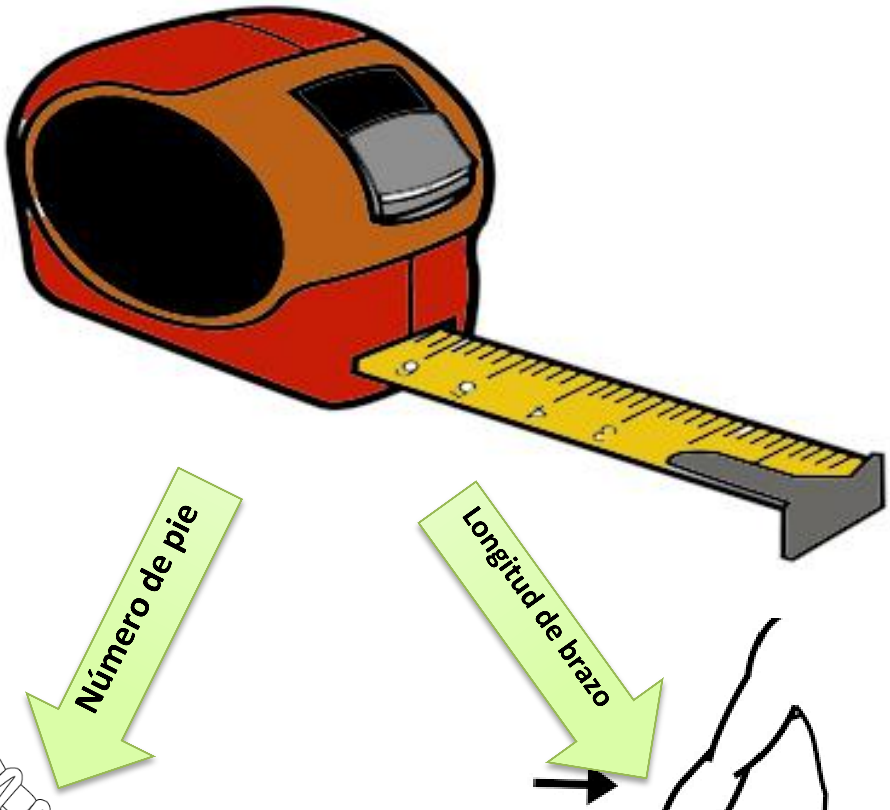
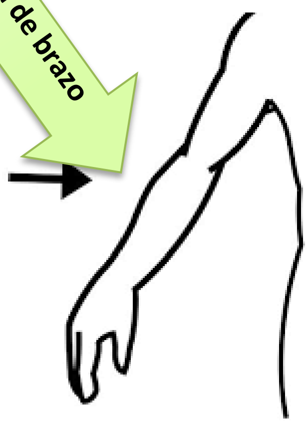
ESTADÍSTICA

INCUBADORA





ÍNDICE

1.- INTRODUCCION

Según Platón, "el comienzo es la parte más importante de la obra", por ello, haremos una digna introducción a nuestro trabajo.

Somos un grupo de jóvenes de 4º de la ESO del centro educativo Presentación de María de Vitoria-Gasteiz. Nuestra profesora de matemáticas, Luisa Aristimuño, considerando que la estadística fue materia de estudio en el primer trimestre, y que la convocatoria tenía lugar con posterioridad, nos incentivó a participar en este concurso. Nosotros aceptamos gustosamente teniendo en cuenta que así podríamos profundizar, con su colaboración, nuestros conocimientos sobre esta rama de las matemáticas: La Estadística y más concretamente sobre la correlación lineal. La espera se nos ha hecho larga, ya que terminamos los temas de probabilidad y estadística en la primera evaluación y creíamos que la convocatoria iba a ser antes del mes de mayo, pero aún así, conservamos la ilusión.

De esta manera formamos un grupo en el cual, además de repasar lo trabajado en la primera evaluación, estudiamos la correlación estadística, dado que en el temario de cuarto, no se estudia en profundidad y la profesora nos comentó que si hacemos ciencias, no trabajaremos estos temas en Bachiller. Esto hizo que aumentara nuestro interés y ganas por hacer este estudio-trabajo.

Participantes:
- Ander Ruiz de Larrinaga Arroyo
- Anne García Urbina
- Danel Arias Alamo
- Nerea Sánchez Fernández

2.- PROFUNDIZACIÓN

Antes de comenzar el trabajo, repasamos los conocimientos que habíamos visto en clase del cálculo de parámetros estadísticos y su significado y nos informamos bien de qué trata la correlación y de cómo hallarla. Para ello, hicimos uso de internet, de los libros y de algunas exposiciones de la profesora, encontrando así toda la información que necesitábamos. Esto fue lo que averiguamos:

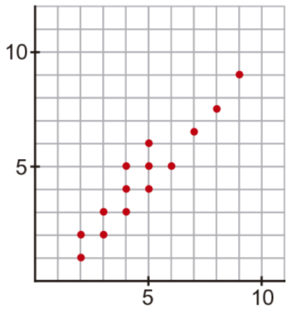
2.1- CORRELACIÓN

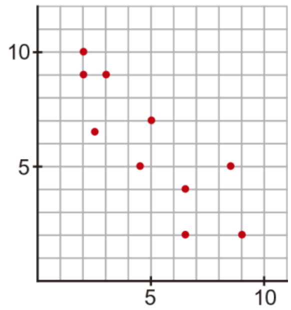
Se considera que dos variables cuantitativas están correlacionadas (también se dice más vulgarmente que están relacionadas) cuando los valores de una de ellas guardan cierta relación con los de la otra: Por ejemplo, si tenemos dos variables diremos que existe correlación si al disminuir los valores de una lo hacen también los de la otra, o al aumentar los valores de una aumentan también los de la otra, siempre y cuando no sea casual.

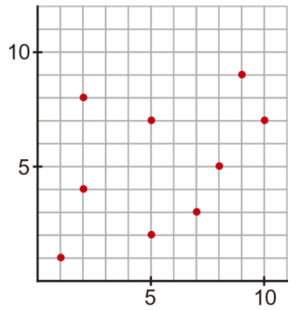
Centramos nuestro estudio en la correlación lineal, es decir, planteamos cuándo las variables se relacionan de forma similar a como lo hacen la X y la Y en una recta. Trazaremos con ayuda de Excel la recta que se llama recta de regresión y nosotros estudiaremos cuándo estas variables están "suficientemente cerca" de esa recta y cuándo no.

La relación entre dos variables cuantitativas queda representada mediante la línea de mejor ajuste, trazada a partir de la nube de puntos. Los principales componentes elementales de una línea de ajuste y, por lo tanto, de una correlación, son la fuerza y el sentido:

- La fuerza, mide el grado en que la línea representa a la nube de puntos: si la nube es estrecha y alargada, se representa por una línea recta, lo que indica que la relación es fuerte; si la nube de puntos tiene una tendencia caótica, la relación es débil. A continuación dibujamos tres ejemplos: en el primero se observa una fuerza mayor que en el segundo y en el tercero la distribución es caótica, es decir, no hay ninguna recta de la que estén cerca casi todos o todos los puntos.







- El sentido mide la variación de los valores de una de las variables con respecto de la otra: si al crecer los valores de una lo hacen los de la otra, la relación es directa o positiva, la recta trazada tendrá pendiente positiva; si al crecer los valores de una disminuyen los de la otra, la relación es inversa o negativa, por lo que la recta trazada tendrá pendiente negativa. En los ejemplos, se ve que la primera es positiva y la segunda negativa. En la tercera también se aprecia que es positiva, pero mucho más débil.

Esta fuerza y el sentido, lo mide el coeficiente de correlación lineal. La fuerza, lo indica el valor absoluto del coeficiente y el sentido lo mide el signo. Describimos esto más detalladamente en el siguiente apartado.

2.2- COEFICIENTE DE CORRELACIÓN LINEAL

El coeficiente de correlación lineal se expresa mediante la letra *r*. Para hallarla, se utiliza la siguiente fórmula: $r = \frac{\sigma_{xy}}{\sigma_x . \sigma_y}$, donde σ_{xy} es la covarianza y σ_x y σ_y las desviaciones típicas de las variables. Es importante tener en cuenta que:

1.-El coeficiente de correlación no varía al hacerlo la escala de medición, es decir, si expresamos la altura en metros o en centímetros el coeficiente de correlación no varía.

2.- El signo del coeficiente de correlación es el mismo que el de la covarianza. Además indica el sentido de la variación: si la covarianza es positiva, la correlación es directa o positiva y si la covarianza es negativa, la correlación es inversa o negativa. Si la covarianza es nula, no existe correlación.

3.- El coeficiente de correlación lineal es un número real comprendido entre menos −1 y 1.

4.- Si el coeficiente de correlación lineal toma valores cercanos al número −1 la correlación es fuerte e inversa o negativa, y será tanto más fuerte cuanto más se aproxime. Análogamente, si el coeficiente de correlación lineal toma valores cercanos al número 1 la correlación es fuerte y directa o positiva, y será tanto más fuerte cuanto más se aproxime r a 1.

5.- Consideraremos que la correlación es fuerte, si el coeficiente de correlación lineal toma valores cuyo valor absoluto es mayor que 0.8. Si r = 1 ó −1, los puntos de la nube están sobre la recta creciente o decreciente, se dice en este caso, que entre ambas variables hay dependencia funcional.

3- PRESENTACIÓN DEL TRABAJO

Antes de comenzar nuestro trabajo, nos informamos de los requisitos que este debía cumplir. Para su elaboración, en vez de usar las técnicas tradicionales (calculadoras, lápiz y papel), aprovechamos nuestros conocimientos sobre las TICs.

Para redactar nuestro proyecto, utilizamos el editor de texto Microsoft Word, y para poder calcular los datos necesarios (dibujar tablas, calcular sumas de columnas, operar con las columnas, hallar medias, desviaciones típicas, covarianza y coeficiente de correlación), utilizamos la herramienta de Microsoft Excel. Ha sido con ayuda de esta herramienta, como hemos obtenido las tablas y diagramas que presentamos en nuestro trabajo. Cuando terminamos de introducir todos los datos y de editarlo, lo convertimos en un PDF.

4.- PROPUESTA DE INCUBADORA

Con nuestro estudio pretendemos saber si el número del calzado y la longitud del brazo derecho, incluyendo la mano, guardan relación, es decir, si estas dos variables estadísticas presentan o no correlación y si dicha correlación es fuerte o débil y directa (o positiva) o indirecta (o negativa).

Esta idea surgió al ver que uno de nuestros compañeros no solo tiene unos brazos extremadamente largos, sino que también un número de calzado que nos parece alto. Propusimos la idea a la profesora como opción de nuestro proyecto y ella contenta, aceptó.

Planteamos quién sería nuestra población a estudiar y vimos que lo más cómodo era preguntar a nuestros compañeros el número de pie y pedirles permiso para medirles el brazo derecho. Muy pronto nos ofrecimos 21 alumnos. Los datos debían estar presentados convenientemente para ser bien leídos e interpretados.

Una tabla presenta las condiciones óptimas para organizar nuestros datos, A continuación, exponemos los datos recogidos:

	NÚMERO DE PIE	LONGITUD DEL BRAZO DERECHO
	X_i	Y_i
Alumno 1	38	69,5
Alumno 2	39	72
Alumno 3	44	74
Alumno 4	39	70,5
Alumno 5	37	68
Alumno 6	38	69
Alumno 7	38	70,5
Alumno 8	38	73
Alumno 9	38	67
Alumno 10	41	72
Alumno 11	40	72
Alumno 12	44	77
Alumno 13	38	68
Alumno 14	38	66
Alumno 15	39	68
Alumno 16	43	71
Alumno 17	42	74
Alumno 18	39	69
Alumno 19	45	73
Alumno 20	40	71,5
Alumno 21	37	65,5

En cada fila aparecen los datos del mismo alumno y en cada columna los de la misma variable: Número de pie y Longitud del brazo derecho.

El primer paso consistió en hacer los cálculos para estudiar la correlación. Con ayuda de Excel, los presentamos:

	NÚMERO DE PIE	LONGITUD DEL BRAZO			
	X_i	Y_i	$X_i.Y_i$	X_i^2	Y_i^2
	38	69,5	2641	1444	4830,25
	39	72	2808	1521	5184
	44	74	3256	1936	5476
	39	70,5	2749,5	1521	4970,25
	37	68	2516	1369	4624
	38	69	2622	1444	4761
	38	70,5	2679	1444	4970,25
	38	73	2774	1444	5329
	38	67	2546	1444	4489
	41	72	2952	1681	5184
	40	72	2880	1600	5184
	44	77	3388	1936	5929
	38	68	2584	1444	4624
	38	66	2508	1444	4356
	39	68	2652	1521	4624
	43	71	3053	1849	5041
	42	74	3108	1764	5476
	39	69	2691	1521	4761
	45	73	3285	2025	5329
	40	71,5	2860	1600	5112,25
	37	65,5	2423,5	1369	4290,25
TOTAL	835	1480,5	58976	33321	104544,25

Calculamos los parámetros necesarios:

$$\overline{X} = \frac{\sum x_i}{n} = \frac{835}{21} = 39{,}7619$$

$$\overline{Y} = \frac{\sum y_i}{n} = \frac{1480{,}5}{21} = 70{,}5$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \overline{X}^2} = \sqrt{\frac{33321}{21} - 39{,}7619^2} = 2{,}3886$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{n} - \overline{Y}^2} = \sqrt{\frac{104544{,}25}{21} - 70{,}5^2} = 2{,}8368$$

$$\sigma_{xy} = \frac{\sum x_i y_i}{21} - \overline{X}.\overline{Y} = \frac{58976}{21} - 39{,}7619.70{,}5 = 5{,}1670$$

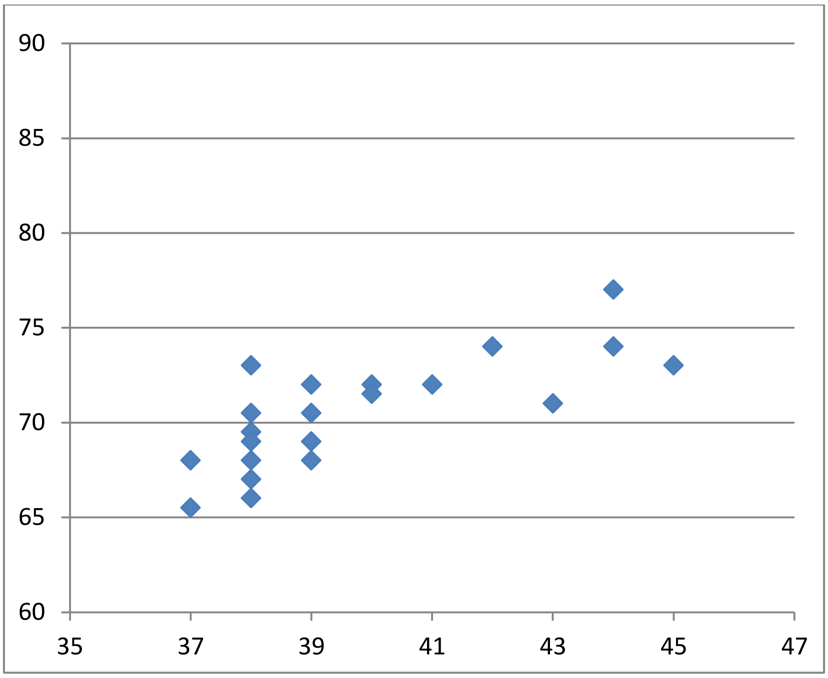
$$r = \frac{\sigma_{xy}}{\sigma_x.\sigma_y} = \frac{5{,}1670}{2{,}3886.2.8368} = 0{,}7625$$

Conclusión:

Hemos obtenido un coeficiente de correlación 0.7625. No está muy lejos de 1 como preveíamos. Claramente, la covarianza es positiva, por lo que la correlación es directa o positiva, es decir, a mayor número de pie, le corresponde mayor longitud del brazo. La correlación nos ha salido fuerte, esto es, hay correlación lineal entre estas dos variables, pero no tan fuerte como esperábamos.

Pasamos a representar los datos de esta variable bidimensional.

La nube de puntos o representación en el plano es la siguiente:



Visualmente, se aprecia una correlación positiva y fuerte.

Nosotros creíamos que el coeficiente de correlación iba a ser mayor y se lo comentamos a la profesora.

La profesora, nos hace la observación de si las medidas las hemos tomado precisas o no. Le contestamos que el punto del hombro, lo hemos elegido sin precisar, sin embargo, el punto de la mano estaba mejor fijado. Con esta observación, decidimos volver a tomar medidas, y concretamos lo siguiente:

Tomamos medida en esta ocasión desde el cambio de juego de los huesos del hombro, hasta el final del dedo corazón de la mano. Teniendo en cuenta este detalle, presentamos los nuevos resultados:

	NÚMERO DE PIE	LONGITUD DEL BRAZO			
	X_i	Y_i	$X_i.Y_i$	X_i^2	Y_i^2
	38	69,5	2641	1444	4830,25
	39	71	2769	1521	5041
	44	77,5	3410	1936	6006,25
	39	70,5	2749,5	1521	4970,25
	37	65	2405	1369	4225
	38	67,5	2565	1444	4556,25
	38	68	2584	1444	4624
	38	70,5	2679	1444	4970,25
	38	65	2470	1444	4225
	41	74	3034	1681	5476
	40	72,5	2900	1600	5256,25
	44	76,5	3366	1936	5852,25
	38	68	2584	1444	4624
	38	64	2432	1444	4096
	39	71	2769	1521	5041
	43	73	3139	1849	5329
	42	74	3108	1764	5476
	39	69	2691	1521	4761
	45	73	3285	2025	5329
	40	71	2840	1600	5041
	37	65,5	2423,5	1369	4290,25
TOTAL	835	1476	58844	33321	104020

Volvemos a calcular los parámetros y en esta ocasión obtenemos:

$$\overline{X} = \frac{\sum x_i}{n} = \frac{835}{21} = 39{,}7619$$

$$\overline{Y} = \frac{\sum y_i}{n} = \frac{1476}{21} = 70{,}2857$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \overline{X}^2} = \sqrt{\frac{33321}{21} - 39{,}7619^2} = 2{,}3886$$

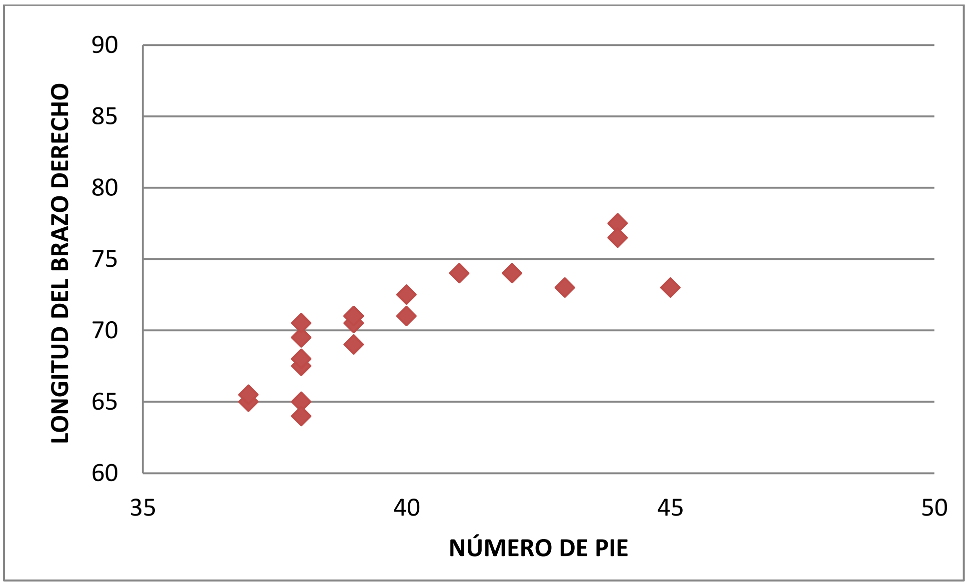
$$\sigma_y = \sqrt{\frac{\sum y_i^2}{n} - \overline{Y}^2} = \sqrt{\frac{104020}{21} - 70{,}2857^2} = 3{,}6405$$

$$\sigma_{xy} = \frac{\sum x_i y_i}{21} - \overline{X}.\overline{Y} = \frac{58844}{21} - 39{,}7619x70{,}2857 = 7{,}4023$$

$$r = \frac{\sigma_{xy}}{\sigma_x . \sigma_y} = \frac{7{,}4023}{2{,}3886x3{,}6405} = 0{,}8513$$

El coeficiente de correlación, está más próximo de 1 que en la primera medición que confesamos haberla hecho con menor precisión y cuidado.

El diagrama de la variable bidimensional es ahora:



En este diagrama, se aprecia una nube de puntos más estrecha que en el anterior, corresponde a una correlación lineal más fuerte como se refleja en los resultados obtenidos. El coeficiente de correlación pasa de ser $r_1 = 0{,}7625$ con la primera medición, a ser $r_2 = 0{,}8513$ con la segunda. Es considerable la diferencia, por lo que habrá que pensar que la exactitud y el cuidado en la toma de datos es muy importante a la hora de hacer cualquier estudio.

5.- CONCLUSIONES

Para finalizar este trabajo, expondremos nuestras conclusiones:

- La longitud del brazo y el número de pie de los 21 alumnos de cuarto de ESO del colegio Presentación de María de Vitoria encuestados son dos variables que presentan correlación lineal fuerte y directa, esto es, a mayor pie de uno de estos alumnos, le corresponde mayor longitud del brazo derecho.
- Es tremendamente importante tomar bien las medidas y determinar con precisión lo que se va a medir, ya que como nos ha ocurrido, si no se determinan de manera minuciosa las variables, puede haber diferencias importantes entre las conclusiones.
- Ha sido importante en nuestro trabajo el conocimiento, uso y manejo de las TICs, pues nos ha permitido presentar el trabajo con más exactitud, comodidad y organización.
- Nos ha parecido una actividad bonita, entretenida e interesante, a la vez que nos ha ayudado a profundizar en la estadística que hemos estudiado en clase, a trabajar con otros compañeros y con la profesora y a ver que esta rama de las Matemáticas tiene su aplicación a la vida.