

Estudio de variables bidimensionales

Unidad 1

Tus profesores te habrán explicado que a veces interesa comparar dos series de datos estadísticos para descubrir influencias o paralelismos entre ellos. Por ejemplo: ¿Existe alguna relación entre el número de películas vistas por Televisión en un mes y el número de horas que cada uno de vosotros dedica al estudio?

Se llaman **variables bidimensionales** a esos datos que se obtienen por parejas, de dos en dos: Número de zapato que calzas y tu estatura, tu edad y tu nivel de colesterol en sangre, etc.

Inicia el programa **LibreOffice.org Calc**. Abre el libro **bidimen1.ods** (Búscalo entre las Herramientas)

Observa que contiene una tabla de datos (*área de datos o tabla principal*), un número llamado R y un gráfico del tipo de **nube de puntos** que representa los números de la tabla.

La tabla está formada por dos columnas, una rotulada con **X** y otra con **Y**. Contiene por tanto **dos variables estadísticas**. La variable X contiene las horas de estudio semanal de un grupo de amigos y la Y las notas de 0 a 10 en un examen.

Elige un par de valores relacionados, por ejemplo **$x=7$ $y=4$** , que corresponde a alguien que estudia 7 horas a la semana y que sacó un 4. Busca qué punto representa a este par en el gráfico. Puedes ir pasando el ratón sobre los puntos de la **nube** para verlo mejor. Haz lo mismo con algún otro.

Intenta lo contrario, buscar un punto en la gráfica y localizar a qué par (x,y) representa. Usa el ratón.

Observa el valor de $R=0,84803$. En este momento ese valor no te dirá nada. Vamos a cambiarlo moviendo los puntos. Para eso debes simplemente cambiar valores de X o de Y. Inténtalo: altera algunos valores de la tabla y observa si R cambia.

El valor del número R, llamado *coeficiente de correlación*, depende de los valores de X e Y

Para entenderlo mejor vas a copiar algunas tablas que contiene el libro **bidim1.ods** un poco más abajo.

Borra el área de datos (tabla principal) con **Supr**

Selecciona la Tabla núm. 1. Usa **Copiar** y **Pegar** para trasladar sus valores a la tabla principal.

Copia el valor de R junto a la Tabla 1. Copia también a su derecha la nube de puntos que ha resultado. Observa que R vale 0,99

Haz lo mismo con la segunda tabla. Ves que R es prácticamente **ceró**. ¿Por qué será tan pequeño? Para responder compara la primera nube con la segunda. Si observas algo interesante escríbelo:

La primera R es mayor porque la nube _____

La segunda R es muy pequeña porque la nube _____

Haz lo mismo con las tablas 3 y 4. ¿Entiendes mejor lo que significa R? Responde:

¿Por qué la tercera tabla tiene un **coeficiente de correlación R** negativo y la primera positivo?

¿Por qué la segunda y la cuarta tienen R casi nula?

¿Qué figura geométrica recuerda la primera nube?

Copia la tercera tabla de nuevo. Recuerda que R vale -0,885.

Intenta ahora algo muy interesante: Cambia algunos valores de X y de Y para conseguir que R se acerque a -1, es decir, que valga -0,95 ó -0,99, etc.

¿Qué le ocurre entonces a la nube? Explícalo:

Ahora intenta que R se acerque a 0, que valga -0,03 ó -0,1, etc.

¿Qué le ocurre entonces a la nube? Explícalo:

Copia ahora la tabla 1, que como sabes, tiene un **coeficiente de correlación R** de 0,99

Intenta que R sea mayor que 1 cambiando valores. ¿Se puede lograr? _____

Resume lo que has aprendido:

Los valores de R no pueden pasar de _____

Si R se acerca a 1 o a -1 los puntos están

Si R se acerca a 0 los puntos están

Inventa una tabla con 10 pares de puntos y que su **coeficiente de correlación R** se acerque a 1

X	Y

Ahora otra en la que se acerque a **cero**

X	Y

Unidad núm. 2

Inicia el programa **LibreOffice.org Calc**. Abre el libro **bidimen2.ods**

Observa que es parecido al libro anterior, pero hay dos novedades interesantes:

- En la nube de puntos aparece una línea de tendencia
- Hay dos celdas para hacer pronósticos

Esto es así porque las nubes de puntos, cuando tienen la forma adecuada, sirven para pronosticar **qué pasaría** con los valores de X que no están en las tablas. Por ejemplo, en la celda G17, rotulada con **Para X =** escribe un valor que no esté en las tablas, por ejemplo 12 y obtendrás debajo un pronóstico para el valor correspondiente de Y: 35,2667 si la tendencia de la tabla se mantiene.

Los pronósticos se calculan siguiendo la línea de tendencia que ves en el gráfico, llamada **recta de regresión**. Esta recta es la que mejor se adapta a la nube.

¿Crees que con la tabla del ejemplo hay un buen ajuste entre nube y recta? _____

¿Influye el valor de R? Para responder a esto cambia algunos valores de Y de forma que ***sus puntos en la nube se alejen de la recta***. ¿Cómo cambia entonces el valor de R? _____

Ahora intenta cambiar valores para que se acerquen a la recta. ¿Qué le ocurre a R?

Copia en el área de datos las tablas que hay más abajo. Observa qué relación hay entre el ajuste de la recta con la nube de puntos y el valor de R.

Resume:

Si los puntos se acercan a la recta de regresión el coeficiente de correlación R se acerca a _____ y si se alejan se acerca a _____

¿Qué ocurre con los pronósticos? Si la R se acerca a 1, ¿tendrán menos errores que si la R es pequeña? Para verlo vas copiando las tablas de nuevo en el área de datos y escribe en los pronósticos valores **de la tabla** y compara su verdadero valor con el pronóstico que te daría la recta de regresión. Rellena la tabla de abajo con lo que has experimentado:

	Valor de X	Valor de Y	Pronóstico	Error
Tabla núm. 1				
Tabla núm. 1				
Tabla núm. 2				

Tabla núm. 2				
Tabla núm. 3				
Tabla núm. 3				
Tabla núm. 4				
Tabla núm. 4				

Habrás descubierto que si R se acerca a 1 o a -1 los errores son más _____ y si se acerca a 0 son más _____

Para verlo mejor en conjunto abre el libro **bidimen3.ods**.

Podrás observar que este libro contiene pronósticos para toda la tabla y los errores cometidos.

Mira en el gráfico una novedad, y es que te aparece la ecuación de la recta de regresión, la que sirve para pronosticar. Es una ecuación de primer grado, del tipo $y = ax+b$. También figura R^2 , que es el cuadrado del coeficiente de correlación, que por ahora no vamos a analizar. Por si tienes curiosidad, es el porcentaje de variabilidad que es explicado por el modelo.

En el modelo **bidimen4.ods** puedes consultar todos los cálculos.