

Para ampliar 7

Estimadores en la regresión y correlación

Este tema se ha considerado siempre como de nivel superior, pero existen situaciones reales que demandan la estimación de parámetros en el modelo de la regresión. Por esto se incluye en estos temas de carácter práctico. Su justificación teórica es más compleja, pero se puede consultar en manuales específicos.

Un ejemplo de la necesidad de estas estimaciones es el hecho de que muchos docentes calculan coeficientes de correlación, pero cuando obtienen valores de tipo medio, como 0,55, 0,62, -0,4 no saben interpretar si estos valores indican que existe correlación significativa.

Usaremos un ejemplo práctico para explicar algunas técnicas:

Cuestión - Ejemplo

¿Es significativa la correlación que encuentro?

Supongamos que un orientador escolar está tratando el tema de la paz con un grupo de alumnos y alumnas. Les ha pasado una prueba en la que la puntuación depende más de los conocimientos previos, y otra, más inclinada a las actitudes. Llamémoslas A y B. La prueba A puntúa entre 0 y 20 y la B entre 0 y 10. Los resultados, ordenados según las puntuaciones en A han sido los siguientes:

| | | | | | | | | | |
|-------------|----|----|----|----|----|----|----|----|----|
| Prueba A 0 | 2 | 2 | 4 | 5 | 5 | 5 | 7 | 8 | 8 |
| Prueba B 2 | 1 | 1 | 6 | 3 | 7 | 2 | 4 | 6 | 6 |
| Prueba A 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 11 | 12 |
| Prueba B 8 | 5 | 2 | 4 | 9 | 3 | 6 | 3 | 7 | 5 |
| Prueba A 12 | 13 | 14 | 14 | 14 | 15 | 15 | 17 | 17 | 18 |
| Prueba B 0 | 9 | 7 | 7 | 6 | 8 | 7 | 9 | 6 | 8 |

¿Es significativamente distinta de cero la correlación entre los resultados de las dos pruebas?

En caso afirmativo, ¿se pueden ajustar estos datos a un modelo lineal $B = a + bA$?

Procedimiento práctico

En primer lugar, mediante una hoja de cálculo, por ejemplo [regresion.ods](#) (tema 5) , calculamos la estimación de los parámetros a, b y r. Copia los datos en dos columnas de la zona de cálculos.

| Columna de la variable X | Columna para la Y | Estimación lineal |
|--------------------------|-------------------|-------------------|
| 0 | 2 | 2,08 |
| 2 | 1 | 2,72 |
| 2 | 1 | 2,72 |
| 4 | 6 | 3,35 |
| 5 | 3 | 3,67 |
| 5 | 7 | 3,67 |
| 5 | 2 | 3,67 |
| 7 | 4 | 4,31 |
| 8 | 6 | 4,63 |
| 8 | 6 | 4,63 |
| 8 | 8 | 4,63 |
| 9 | 5 | 4,95 |

Los resultados son:

Coefficiente de correlación $r = 0,575$

Pendiente estimada $B = 0,319$

Ordenada en el origen $A = 2,080$

| | | | |
|-------------|-------|-----------------------------|-------------------------------------|
| Media de X | 9,900 | Coefficiente de correlación | 0,575 |
| Media de Y | 5,233 | Error de estimación | 2,173 |
| Des. Tip. X | 4,629 | Pronóstico individual | |
| Des. Tip. Y | 2,565 | Escribe un valor de X | <input type="text" value="26,000"/> |
| Regresión | | Su Y estimada será | <input type="text" value="10,361"/> |
| Pendiente | 0,319 | | |
| Ordenada | 2,080 | | |

¿Son significativos estos valores?



En el [resumen teórico](#) puedes repasar los estimadores y las distribuciones muestrales de los parámetros en la Regresión y Correlación.

Coefficiente de correlación

El coeficiente obtenido 0,575 nos plantea dos cuestiones. La primera es si es significativamente distinto de cero. Para ello puedes usar la misma herramienta

Si has mantenido escritos los datos, pasa a la hoja **Estimación**. Junto al valor 0,575 figura el valor de T obtenido de la fórmula contenida en la teoría

$$T = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

con un resultado de 3,716. Este valor es muy alto, y si lees la probabilidad de su aparición (p-valor=0,001) te darás cuenta de que es muy improbable una correlación así por puro azar, luego podemos aceptar que

El coeficiente de correlación 0,575 es significativamente distinto de cero, luego podemos concluir que verdaderamente existe una correlación entre las dos pruebas por el orientador.

La segunda cuestión es la de estimar ese mismo coeficiente en la población. El intervalo de confianza requiere para su cálculo la transformada de Fisher, explicada en la teoría. Aquí no profundizaremos más.

Coeficientes A y B

En los casos como el anterior, en los que existe correlación significativa, se pueden estimar mediante intervalos los coeficientes a y b mediante sus estimadores A y B, que en el ejemplo anterior presentaban los valores $B=0,319$ y $A=2,080$. En la [teoría](#) puedes consultar sus distribuciones muestrales.

En la herramienta regresion.ods se utilizan estas distribuciones muestrales para construir intervalos de confianza, resultando:

Para A: (0,16 , 4)

Para B: (0,14 , 0,49)

En estos temas no profundizaremos más en estas técnicas. Tan sólo se desea presentar sus posibilidades.