

REGRESIÓN LINEAL

#Act 1. Si pretendiésemos explicar un suceso y/o fenómeno acontecido en el pasado ¿Podemos inferir la respuesta asociada a dichos eventos en base a los restos materiales presentes?

No podemos ya que los materiales que están presentes podrían mostrar pistas o indicios de eventos ocurridos en el pasado.

#Act 2. Haciendo referencia al análisis de correlación lineal de Pearson, ¿establece este algún tipo de relación causa-efecto de una variable sobre otra(s)?

No, el análisis de correlación lineal de Pearson no llega a establecer una relación basada en el causa-efecto de las variables, únicamente nos indica la fuerza y la dirección.

#Act 3. Define causalidad. Exponga algún ejemplo.

Aquella relación que tiene un evento al producir un efecto. Por ejemplo, cae un rayo (causa) y se produce un fuego (efecto).

#Act 4. ¿Podrías mencionar los parámetros involucrados en la ecuación de regresión lineal?

Los parámetros son: la pendiente, B_1 ($B=\text{beta}$); y el intercepto, B_0 .

#Act 5. En un plano cartesiano, si afirmo que el eje 'x' también se denomina eje de ordenadas, ¿estoy en lo cierto?

No, ya que el eje X se denomina eje de abscisas y el eje Y se denomina eje de ordenadas.

#Act 6. ¿Sabrías diferenciar entre recta de regresión y plano de regresión?

La recta de regresión es un modelo de regresión lineal que se ajusta a una dimensión concreta, a diferencia del plano de regresión se ajusta a dos dimensiones.

#Act 7. ¿Cuáles son los supuestos (o hipótesis) del análisis de regresión lineal?

Linealidad, independencia, homocedasticidad, normalidad y ausencia de multicolinealidad.

#Act 8. Dados los siguientes datos, calcula la recta de regresión que mejor se adapte a nuestra nube de puntos siendo "cuentas" la variable dependiente o de respuesta y "distancia" la variable independiente o explicativa.

```R

#Datos

```
cuentas <- c(110,2,6,98,40,94,31,5,8,10)
distancia <- c(1.1, 100.2, 90.3, 5.4, 57.5, 6.6, 34.7, 65.8, 57.9, 86.1)
```

```
#recta de regresión
modelo <- lm(cuentas ~ distancia)
summary(modelo)
....
```

#Act 9. Serías capaz de interpretar el significado de los parámetros de la ecuación de regresión?

La pendiente nos indica el cambio promedio en la variable dependiente por cada unidad de cambio en la variable independiente, mientras que el intercepto indica el valor esperado de la variable dependiente cuando la variable independiente equivale a 0.

#Act 10. ¿Qué implicaciones conlleva obtener un intercepto con valor '0'?

Obtener un intercepto con valor 0, supone que no hay efecto de la variable independiente sobre la variable dependiente, ya que la anterior es 0.

#Act 11. ¿Qué ponderación lleva a cabo el análisis de regresión lineal para calcular los valores de los parámetros que configuran la recta de regresión?

Emplea una ponderación de mínimos cuadrados para calcular los valores de los parámetros de la recta.

#Act 12. ¿Cuál sería el error asociado a mi modelo en la estimación del número de cuentas para un yacimiento que se encuentra a 1.1 km de la mina?

Este error recibe el nombre de error estándar residual del modelo.

#Act 13. ¿Cómo calcularías los residuos del modelo dado los siguientes datos?

Los residuos se calculan restando las predicciones del modelo de los valores observados.

#Act 14. Con los datos residuales, verifica si se cumple (o no) el supuesto de normalidad.

```
```R
```

```
# Datos
```

```
```R
```

```
#Datos
```

```
cuentas <- c(110,2,6,98,40,94,31,5,8,10)
distancia <- c(1.1, 100.2, 90.3, 5.4, 57.5, 6.6, 34.7, 65.8, 57.9, 86.1)
```

```
#recta de regresión
modelo <- lm(cuentas ~ distancia)
```

```
summary(modelo)
```

```
....
```

```
Verifica la normalidad de los residuos
```

```
shapiro.test(residuos)
```

```
...
```

#Act 15. ¿Que 2 de conjuntos (de datos) se han de emplear en la modelización lineal?  
¿Cómo llevarías a cabo la preparación de estos?

Se emplean dos conjuntos de datos: uno para el entrenamiento del modelo y otro para la validación. Para prepararlos, se dividen los datos disponibles en conjuntos de entrenamiento y validación.

#Act 16. Evalúa la capacidad predictiva del modelo implementando una validación cruzada simple.

La capacidad predictiva del modelo se evalúa mediante técnicas de validación cruzada.

#Act 17. Si mis coeficientes de regresión se han calculado con un intervalo de confianza del 95% ¿cuál será la probabilidad de que la correlación lineal entre los coeficientes de regresión y la variable de respuesta o explicada se deba al azar? ¿Y si tengo un nivel de significación (Alpha) de 0.01, con que Intervalo de Confianza he obtenido mis coeficientes de regresión?

La probabilidad de que se deba al azar es 0.05 si el intervalo de confianza es del 95%. Si el nivel de significación es 0.01, el intervalo de confianza será del 99%.

#Act 18. Si las estimaciones arrojadas por mi modelo lineal resultan menos precisas (mayor error) en un determinado rango de valores con respecto a otro, decimos ¿qué hay indicios de homocedasticidad o heterocedasticidad?

Si el modelo resulta menos preciso en un determinado rango de valores, indica la presencia de heterocedasticidad.

#Act 19. ¿Qué medida de precisión estadística nos indica el % de variabilidad explicada de la variable dependiente por nuestro modelo lineal?

El coeficiente de determinación ( $R^2$ ) indica el porcentaje de variabilidad explicada de la variable dependiente por el modelo lineal.

#Act 20. Explica la diferencia entre una observación atípica y una observación que produzca lo que se conoce en estadística como “apalancamiento” del modelo?

Una observación atípica es un valor inusual en los datos, mientras que una observación que produce apalancamiento del modelo es una observación que ejerce una influencia desproporcionada en la estimación de los parámetros del modelo.