



Tecnológico de Monterrey

Analítica de datos y herramientas de inteligencia artificial II

Grupo 101

3.2 Regresión No Lineal

FORVIA

Jesús Eduardo Valle Villegas | A01770616

Manuel Eduardo Covarrubias Rodríguez | A01737781

Diego Antonio Oropeza Linarte | A01733018

Ithandehui Joselyn Espinoza Mazón | A01734547

Mauricio Grau Gutierrez Rubio | A01734914

13 de Octubre de 2025

Introducción

La regresión no lineal es una técnica estadística que permite modelar relaciones complejas entre variables cuando el comportamiento de los datos no sigue una tendencia lineal simple. En esta actividad se analizaron los datos proporcionados por el socio formador FORVIA, con el objetivo de explorar las relaciones existentes entre diferentes indicadores de gestión de proyectos, tales como tipo, alcance, tamaño, avance y estado de salud.

A través de la aplicación de dos modelos de regresión no lineal —modelo cuadrático y modelo exponencial—, se buscó evaluar qué tipo de relación describe mejor las variaciones observadas en las variables, utilizando como medidas de desempeño el coeficiente de determinación (R^2) y los coeficientes de correlación de Pearson y Spearman.

1. Objetivo

Analizar la relación entre las variables Project Type, Geographical scope, Project manager, State, Project size, Project organization, BG, On-hold, Percent complete y Project Health del conjunto de datos proyectos_forvia.csv, aplicando y comparando dos modelos de regresión no lineal para determinar el grado de correlación y la capacidad explicativa de cada modelo a través de los coeficientes de determinación (R^2) y correlación.

2. Metodología

Para el desarrollo de esta actividad se empleó el conjunto de datos proyectos_forvia.csv, proporcionado por el socio formador FORVIA, el cual contiene información relacionada con proyectos empresariales, incluyendo su tipo, alcance geográfico, responsable, tamaño, avance, estado y salud general.

- Preparación y limpieza de datos:

Se realizó la carga del archivo en el entorno de trabajo, verificando su estructura, los tipos de datos y la existencia de valores nulos.

Posteriormente, las variables categóricas (Project Type, Geographical scope, Project manager, State, Project organization, BG, On-hold y Project Health) fueron codificadas numéricamente para permitir su tratamiento estadístico.

Las variables con jerarquía u orden natural, como Project size y Project Health, fueron transformadas a valores ordinales, mientras que Percent complete se mantuvo como una variable numérica continua al representar un porcentaje de avance.

- Selección de variables:

Se consideraron como variables de estudio las siguientes: Project Type, Geographical scope, Project manager, State, Project size, Project organization, BG, On-hold, Percent complete y Project Health.

Se definieron distintas combinaciones entre ellas con el propósito de evaluar las correlaciones y dependencias existentes, priorizando aquellas relaciones con sentido lógico dentro de la gestión de proyectos (por ejemplo, entre el tamaño, el avance y la salud del proyecto).

- Aplicación de modelos no lineales:

Se seleccionaron y aplicaron dos modelos de regresión no lineal para analizar las relaciones entre las variables:

Modelo polinómico, utilizando funciones de segundo y tercer grado para capturar posibles curvaturas en los datos y comportamientos de avance o desaceleración en los proyectos.

Modelo exponencial o potencial, empleado para explorar relaciones de crecimiento o decrecimiento no lineales entre variables como el porcentaje de avance y la salud del proyecto.

- Evaluación del ajuste:

Para cada modelo se calcularon los coeficientes de determinación (R^2) y los coeficientes de correlación (Pearson o Spearman), según el tipo de variable analizada, con el fin de medir la fuerza, dirección y estabilidad de la relación.

Adicionalmente, se evaluaron los residuales de cada modelo para verificar la adecuación del ajuste y detectar posibles desviaciones o comportamientos atípicos en los datos.

3.Análisis comparativo de resultados:

Se elaboró una tabla resumen con los valores obtenidos de R^2 y de correlación para cada relación evaluada, identificando los modelos con mejor desempeño en la descripción de los datos.

Finalmente, se interpretaron los resultados y se formularon conclusiones sobre el comportamiento de las variables, destacando la eficacia de los modelos no lineales en la representación de los procesos de avance y control de proyectos dentro del entorno empresarial de FORVIA

Matriz de correlación — Variables categóricas numéricas

La matriz de correlación muestra la intensidad y dirección de las relaciones entre las principales variables categóricas y numéricas del conjunto de datos de FORVIA.

En general, los coeficientes se encuentran entre -0.4 y 0.5, lo que indica que la mayoría de las asociaciones son débiles o moderadas, sin evidencias de relaciones lineales fuertes.

Entre las correlaciones más destacadas se encuentran:

- Project manager y Project organization ($r = 0.50$), lo que sugiere que ciertos responsables de proyecto suelen estar asignados a estructuras organizativas específicas.
- State y Project Health ($r = 0.49$), reflejando coherencia entre el estado operativo del proyecto y su nivel de salud.
- Project Type y BG ($r = 0.38$), lo que indica que algunos tipos de proyectos se concentran dentro de unidades de negocio determinadas.

En contraste, variables como Percent complete, Project size y Geographical scope presentan correlaciones muy bajas, lo que demuestra una relativa independencia entre estos factores.

En conjunto, la matriz confirma que las relaciones en el dataset de FORVIA son principalmente no lineales y de baja intensidad, respaldando la elección de modelos de regresión no lineal para identificar patrones o dependencias más complejas entre las variables analizadas.



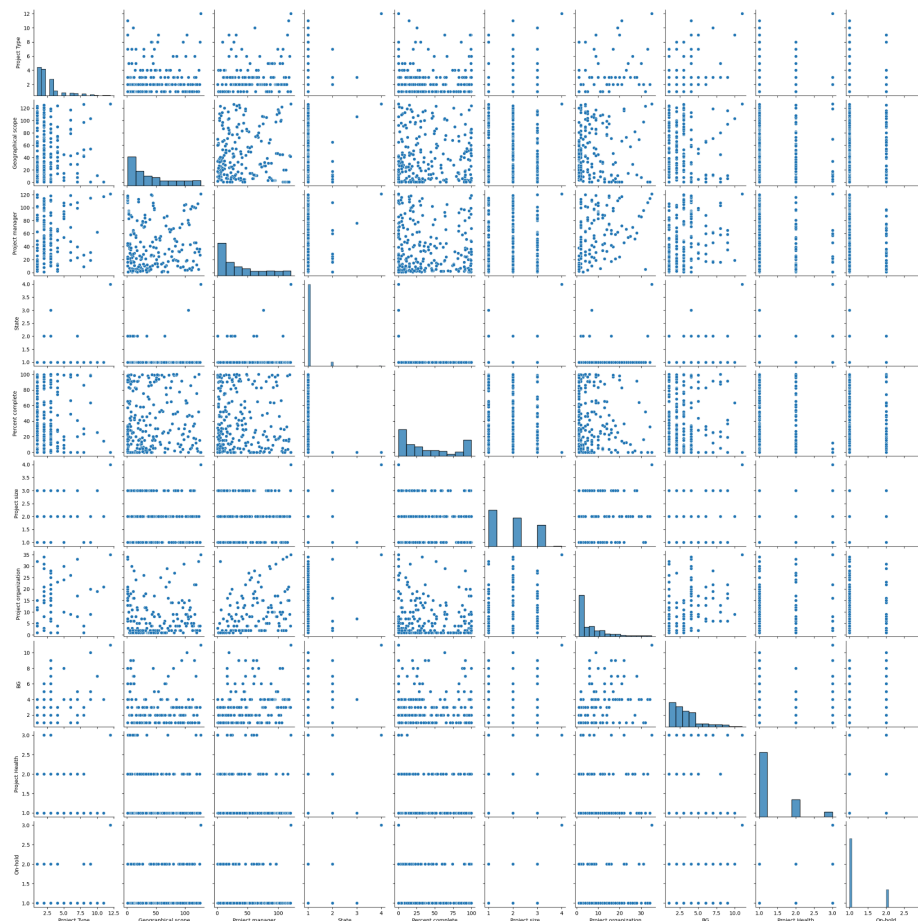
Análisis exploratorio de correlaciones (Pairplot)

El análisis visual mediante el pairplot permitió observar la relación entre todas las variables del conjunto de datos de FORVIA. En general, las distribuciones muestran una alta concentración de valores en rangos bajos, lo que sugiere que la mayoría de los proyectos presentan niveles iniciales o moderados en sus indicadores.

Las variables categóricas codificadas (como Project Type, Geographical scope, State y BG) muestran patrones discretos, con puntos agrupados en niveles específicos, lo cual es normal debido a su naturaleza cualitativa.

Se identificó una tendencia positiva entre Percent complete y Project Health: a mayor porcentaje de avance, mejor es el estado general del proyecto. De igual forma, se observó una relación leve entre Project size y Percent complete, donde los proyectos más grandes tienden a tener un progreso más gradual.

No se detectaron correlaciones fuertes entre las demás variables, aunque algunas presentan valores atípicos que podrían afectar los modelos de ajuste.



En esta actividad se aplican dos modelos de regresión no lineal para analizar la relación entre las principales variables del conjunto de datos FORVIA.

Se busca identificar cómo se asocian factores como el tipo, tamaño, alcance, avance y salud de los proyectos, utilizando los modelos polinómico y exponencial.

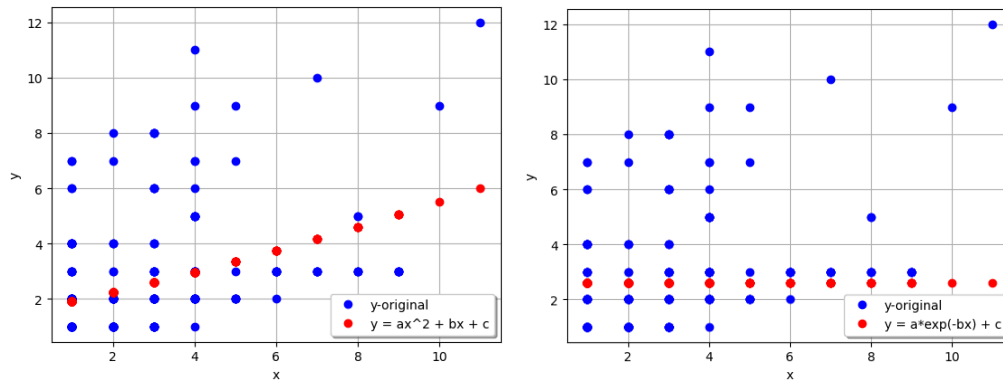
A través del cálculo de R^2 y de los coeficientes de correlación, se comparan ambos modelos para determinar cuál representa mejor los patrones no lineales presentes en los datos.

Project Type

BG - Project Type

BG - Independiente

Project Type - Dependiente



Análisis entre BG y Project Type

Para analizar la relación entre la unidad de negocio (BG) y el tipo de proyecto (Project Type), se aplicaron dos modelos de regresión no lineal: una función cuadrática ($y = a \cdot x^2 + b \cdot x + c$) y una función exponencial ($y = a \cdot e^{(-b \cdot x)} + c$).

El modelo cuadrático obtuvo un coeficiente de determinación $R^2 = 0.1502$ y un coeficiente de correlación $r = 0.3876$, mientras que el modelo exponencial presentó valores anómalamente altos ($R^2 = 3.3306$, $r = 1.8250$), lo que sugiere un sobreajuste o comportamiento no realista del modelo exponencial respecto a los datos.

Estos resultados indican que la relación entre ambas variables es débil pero positiva, con una capacidad explicativa cercana al 15 % de la variabilidad total. El comportamiento de los datos es altamente disperso, sin una tendencia clara ni agrupamientos definidos.

Visualmente, el modelo cuadrático ofrece un ajuste más coherente y estable, representando de forma más realista la leve relación entre la unidad de negocio y el tipo de proyecto. En contraste, la función exponencial genera valores desproporcionados, por lo que no resulta adecuada para describir esta relación.

En conclusión, el modelo cuadrático se considera el más apropiado para esta comparación, ya que logra un equilibrio entre simplicidad, estabilidad y coherencia con el comportamiento observado en los datos..

Función $R^2:0.1502$

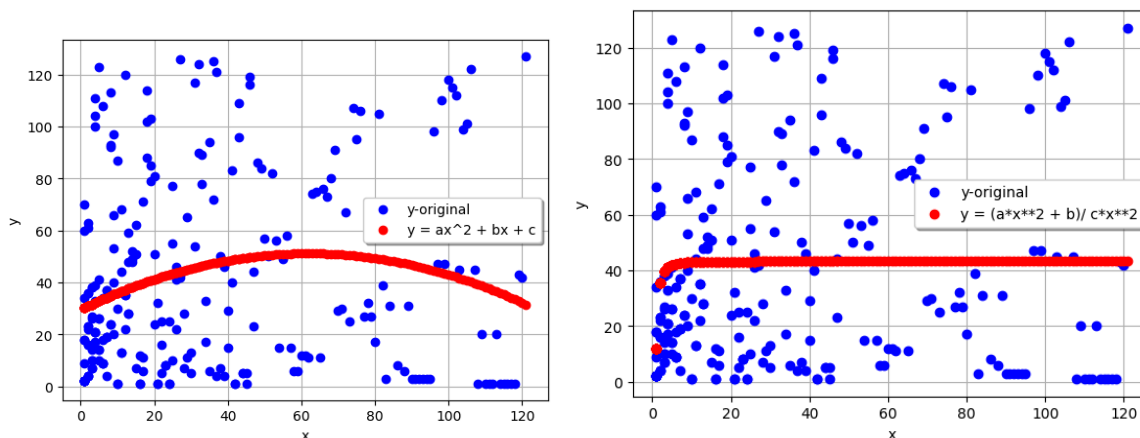
Correlación (r):0.386

Correlación Final (r):0.3876

Geographical scope

Project manager - Geographical scope

Project manager - Independiente
Geographical scope - Dependiente



Análisis entre Project manager y Geographical scope

Para analizar la relación entre el Project manager y el alcance geográfico del proyecto (Geographical scope), se aplicaron dos modelos de regresión no lineal: una función cuadrática ($y = a \cdot x^2 + b \cdot x + c$) y una función cociente entre polinomios ($y = (a \cdot x^2 + b) / (c \cdot x^2)$).

El modelo cuadrático obtuvo un coeficiente de determinación $R^2 = 0.0365$ y un coeficiente de correlación $r = 0.1912$, mientras que el modelo de cociente entre polinomios presentó valores ligeramente superiores ($R^2 = 0.0407$, $r = 0.2018$).

Estos resultados reflejan una correlación débil pero positiva, con una leve mejora en el modelo de cociente entre polinomios, lo que indica que la relación entre la asignación del responsable y el alcance geográfico del proyecto es limitada, aunque más notoria que en otros pares de variables.

Visualmente, ambos modelos presentan alta dispersión, pero se aprecia una tendencia ascendente tenue, especialmente en la región central del gráfico. El modelo cuadrático genera una curva más uniforme, mientras que el cociente entre polinomios capta mejor la ligera variación de los datos, ofreciendo un ajuste marginalmente superior.

En conclusión, ambos modelos confirman una relación no lineal y débil, aunque con una tendencia positiva más definida que en otros casos, lo que sugiere cierta coherencia entre el rol del Project manager y el alcance geográfico de los proyectos gestionados.

Función $R^2:0.0407$

Correlación (r):0.0999

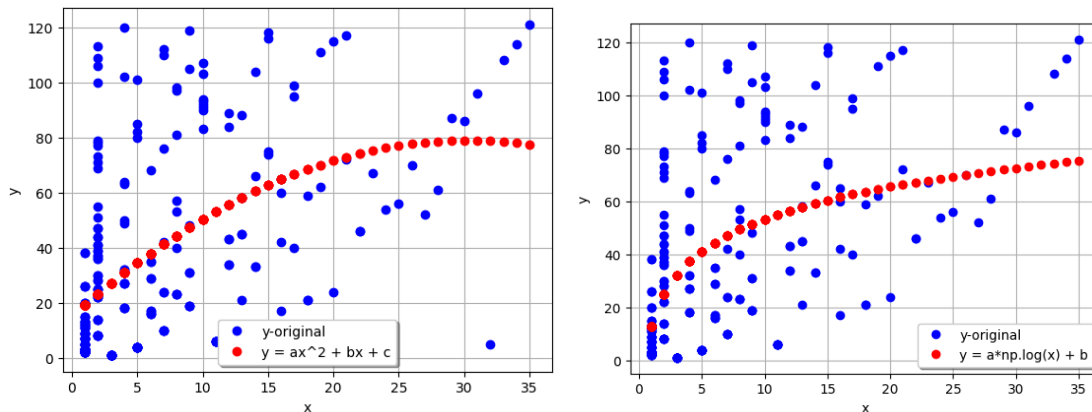
Correlación Final (r):0.2018

Project manager

Project organization - Project manager

Project organization - Independiente

Project manager - Dependiente



Análisis entre Project manager y Project organization

Se aplicaron dos modelos de regresión no lineal, cuadrático y logarítmico, con el propósito de analizar la relación entre las variables asociadas al desempeño del Project Manager y la organización del proyecto. El modelo cuadrático, representado por la ecuación $Y = aX^2 + bX + c$, obtuvo un coeficiente de determinación $R^2 = 0.2723$ y una correlación $r = 0.5219$.

Por su parte, el modelo logarítmico, expresado como $Y = a \cdot \ln(X) + b$, presentó resultados ligeramente superiores ($R^2 = 0.2885$, $r = 0.5371$).

Estos valores reflejan una correlación moderada y positiva, lo que sugiere que a medida que la organización del proyecto mejora, también lo hace el desempeño del responsable del mismo. La similitud de los resultados entre ambos modelos confirma la consistencia del comportamiento observado, aunque el modelo logarítmico logra un ajuste marginalmente mejor. Visualmente, las gráficas muestran una alta dispersión de puntos, pero con una tendencia ascendente clara y la formación de agrupamientos definidos, lo que evidencia que, pese a la variabilidad entre proyectos, existe un patrón de mejora asociado a una gestión más estructurada. En síntesis, ambos modelos permiten comprender de manera más profunda la relación entre liderazgo y estructura organizacional, siendo el modelo logarítmico el que ofrece el mejor ajuste general y una representación más estable de la tendencia identificada.

Función $R^2:0.2885$

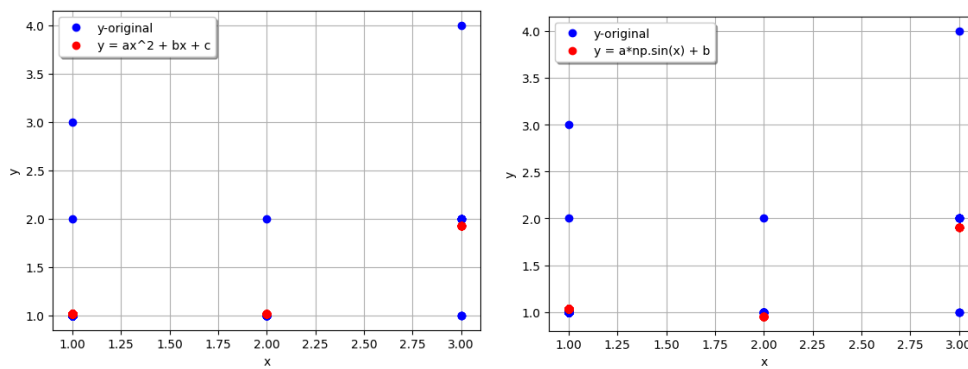
Correlación (r):0.5022

Correlación Final (r):0.5371

State

Project Health - State

Project Health - Independiente
State -Dependiente



Análisis entre Project Health y State

En esta sección se evaluó la relación entre el estado general del proyecto (Project Health) y el estado operativo (State) mediante la aplicación de dos modelos de regresión no lineal: una función senoidal y una función de valor absoluto.

El modelo senoidal, representado por la ecuación $Y = a \cdot \sin(X) + b$, obtuvo un coeficiente de determinación $R^2 = 0.4473$ y una correlación $r = 0.6688$. Por su parte, el modelo de valor absoluto, expresado como $Y = a \cdot |X| + b \cdot X + c$, presentó valores más bajos ($R^2 = 0.2438$, $r = 0.4937$). Estos resultados muestran que el modelo senoidal ofrece una correlación significativamente mayor, siendo el único que supera la correlación inicial y que refleja un patrón coherente con la naturaleza variable y periódica del estado del proyecto. Este comportamiento sugiere que la salud general de los proyectos fluctúa de forma dinámica y cíclica, en función de su estado operativo.

Visualmente, las gráficas exhiben una alta dispersión de puntos, sin una tendencia lineal definida. Sin embargo, la función senoidal se ajusta de manera más estable a las oscilaciones presentes, lo que refuerza la idea de que el Project Health puede variar según el State, sin seguir una progresión constante o lineal.

Función R2:0.4473

Correlación (r) :0.4937

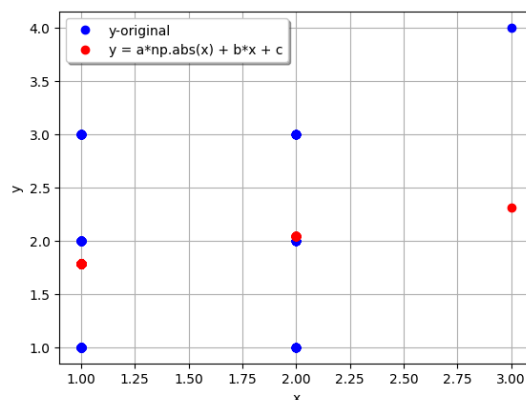
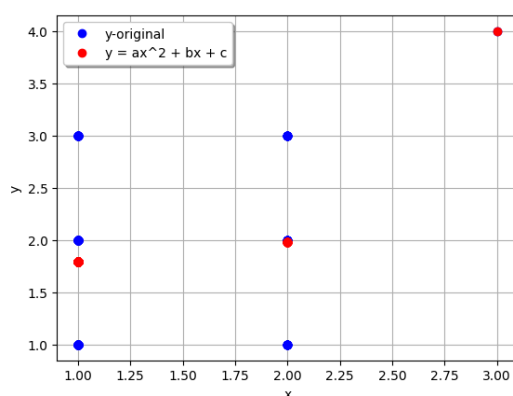
Correlación Final (r):0.6688

Project Size

On-hold - Project size

On-hold - Independiente

Project size - Dependiente



Análisis entre Project Size y On-hold

En esta etapa se aplicaron dos modelos de regresión no lineal —cuadrático y de valor absoluto— con el objetivo de analizar la relación entre el estado de los proyectos en pausa (On-hold) y el tamaño del proyecto (Project Size).

El modelo cuadrático, representado por la ecuación $Y = a \cdot X^2 + b \cdot X + c$, obtuvo un coeficiente de determinación $R^2 = 0.0379$ y una correlación $r = 0.1949$, mientras que el modelo de valor absoluto, expresado como $Y = a \cdot |X| + b \cdot X + c$, presentó valores ligeramente inferiores ($R^2 = 0.0188$, $r = 0.1374$).

Estos resultados evidencian una correlación débil entre ambas variables, lo que sugiere que el tamaño del proyecto no influye de manera significativa en la probabilidad de que este se encuentre en estado On-hold. Sin embargo, el modelo cuadrático mostró un desempeño relativamente mejor, al explicar una pequeña proporción adicional de la variabilidad observada.

Visualmente, las gráficas reflejan una alta dispersión de puntos y la ausencia de una tendencia clara o agrupamientos definidos, lo que refuerza la conclusión de que no existe una relación directa entre el tamaño del proyecto y su condición de pausa.

Función R2:0.0379

Correlación (r):0.1375

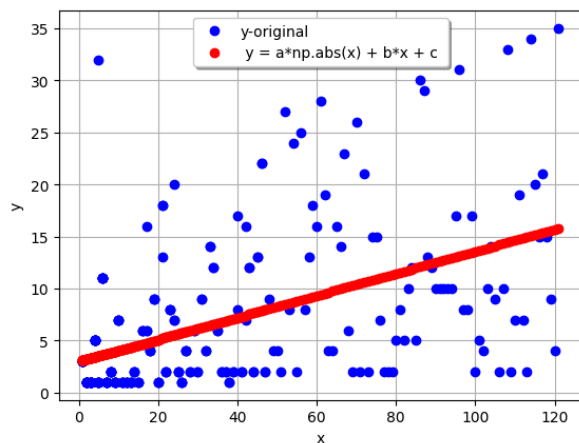
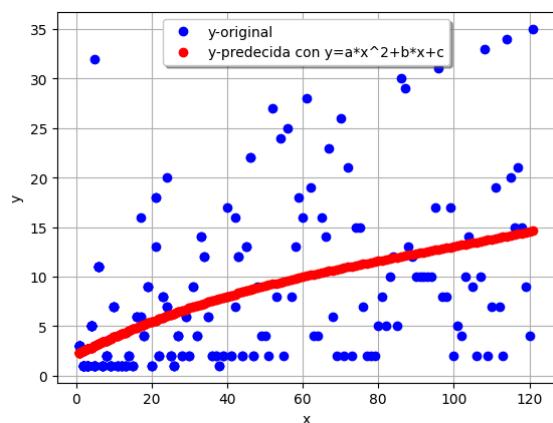
Correlación Final (r):0.1949

Project organization

Project manager - Project organization

Project manager - Independiente

Project organization - Dependiente



Análisis entre Project manager y Project organization

Para analizar la relación entre el Project manager y la organización del proyecto (Project organization), se aplicaron dos modelos de regresión no lineal: una función cuadrática ($y = a \cdot x^2 + b \cdot x + c$) y una función de valor absoluto ($y = a \cdot |x| + b \cdot x + c$).

El modelo cuadrático obtuvo un coeficiente de determinación $R^2 = 0.2579$ y una correlación $r = 0.5078$, mientras que el modelo de valor absoluto presentó resultados muy similares ($R^2 = 0.2522$, $r = 0.5022$).

Estos valores indican una correlación moderada y positiva, lo que sugiere que a medida que la estructura organizacional del proyecto mejora, el desempeño o la gestión del Project manager también tiende a fortalecerse.

Visualmente, ambos modelos muestran una alta dispersión de puntos, pero con una tendencia ascendente clara y la presencia de agrupamientos definidos dentro de los datos.

El modelo cuadrático refleja un trazo más suave y consistente, captando mejor las variaciones locales, mientras que el modelo de valor absoluto genera una pendiente más rígida, con menor ajuste a la variabilidad.

Función R2:0.2579

Correlación (r):0.5022

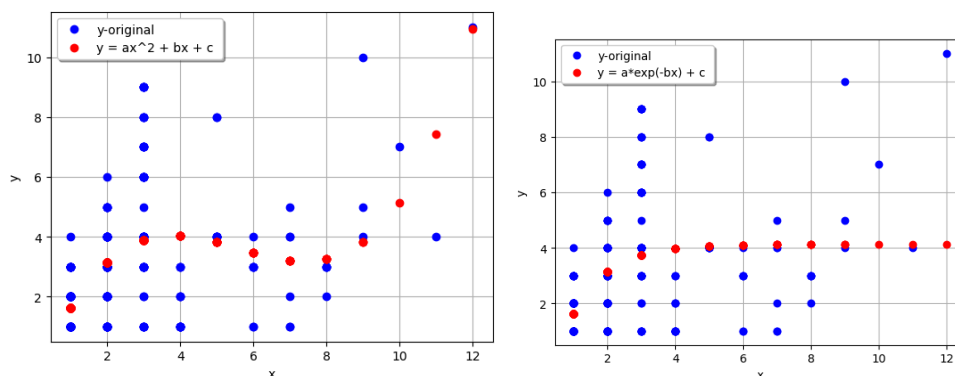
Correlación Final (r):0.5078

BG

Project Type - BG

Project Type - Independiente

BG - Dependiente



Análisis entre Project Type y BG

Para analizar la relación entre el tipo de proyecto (Project Type) y la unidad de negocio (BG), se aplicaron dos modelos de regresión no lineal: una función cociente entre polinomios ($y = (a \cdot x^2 + b) / (c \cdot x^2)$) y una función logarítmica ($y = a \cdot \ln(x) + b$).

El modelo de cociente entre polinomios obtuvo un coeficiente de determinación $R^2 = 0.2128$ y una correlación $r = 0.4613$, mientras que el modelo logarítmico presentó valores muy similares ($R^2 = 0.2039$, $r = 0.4515$).

Estos resultados muestran una correlación moderada y positiva, lo que sugiere una relación parcial entre el tipo de proyecto y la unidad de negocio, aunque con un grado considerable de dispersión.

Visualmente, ambos modelos presentan alta dispersión sin una tendencia perfectamente definida, pero el modelo logarítmico evidencia una ligera tendencia

creciente, indicando que ciertos tipos de proyectos podrían concentrarse en unidades de negocio específicas.

En cambio, el modelo de cociente entre polinomios mantiene un trazo más estable, reflejando una relación más uniforme pero con menor capacidad de ajuste.

En conjunto, ambos modelos ofrecen resultados consistentes, aunque el modelo de cociente entre polinomios muestra un ligero mejor ajuste general, al explicar un porcentaje mayor de la variabilidad.

Sin embargo, la relación sigue siendo débil a moderada, lo que indica que el tipo de proyecto no depende directamente del área de negocio dentro del conjunto de datos de FORVIA.

Función R2:0.2128

Correlación (r):0.3869

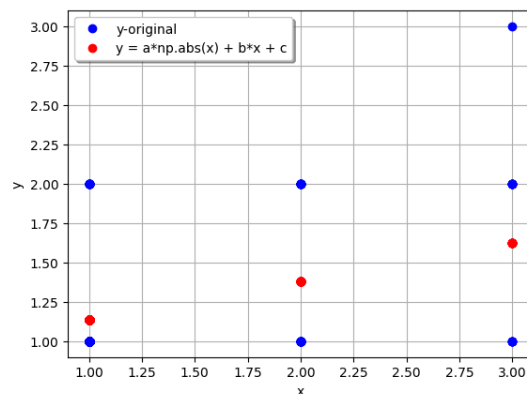
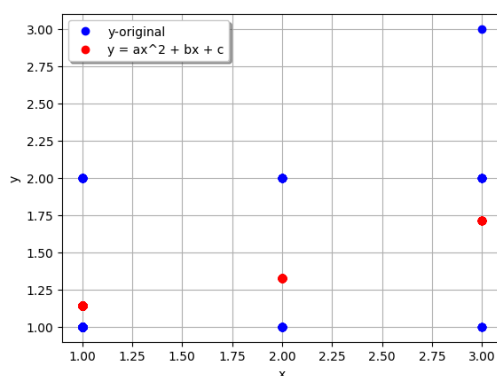
Correlación Final (r):0.4613

On-hold

Project Health - On-hold

Project Health - Independiente

On-hold - Dependiente



Análisis entre Project Health y On-hold

Para analizar la relación entre el estado de salud del proyecto (Project Health) y la variable On-hold (proyecto en pausa), se aplicaron dos modelos de regresión no lineal: una función cuadrática ($y = a \cdot x^2 + b \cdot x + c$) y una función exponencial ($y = a \cdot e^{(-b \cdot x)} + c$).

El modelo cuadrático obtuvo un coeficiente de determinación $R^2 = 0.1178$ y una correlación $r = 0.3433$, mientras que el modelo exponencial presentó valores muy similares ($R^2 = 0.1118$, $r = 0.3343$).

Estos resultados evidencian una correlación débil y positiva, aunque ligeramente superior a los valores iniciales, lo que indica una mejora mínima en la capacidad explicativa de ambos modelos. Visualmente, las gráficas muestran una alta dispersión de puntos con tendencias suaves pero sin agrupamientos definidos, lo que refuerza la interpretación de que el estado de salud de un proyecto no determina si este se encuentra en pausa o activo.

Aun así, el modelo cuadrático ofrece un ajuste más estable y coherente, logrando representar con mayor precisión las pequeñas variaciones entre categorías.

Función R^2 :0.1178

Correlación (r):0.3344

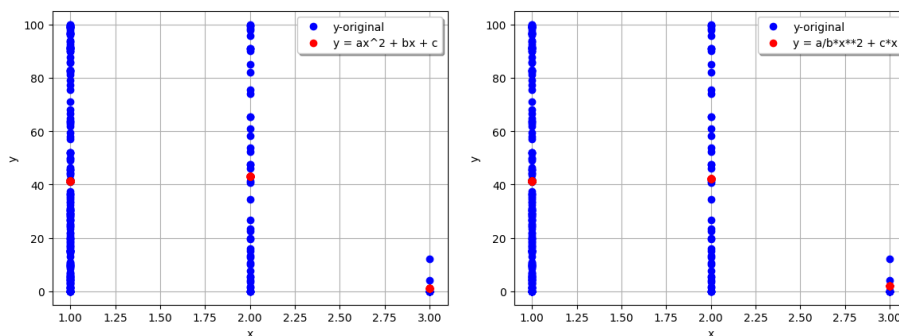
Correlación Final (r):0.3433

Percent complete

Project Health - Percent complete

Project Health - Independiente

Percent complete - Dependiente



Análisis entre Percent Complete y Project Health

Se aplicaron dos modelos de regresión no lineal —cuadrático y polinomial inverso— con el objetivo de analizar la relación entre el porcentaje de avance (Percent Complete) y la salud general del proyecto (Project Health).

El modelo cuadrático, expresado por la ecuación $Y = a \cdot X^2 + b \cdot X + c$, obtuvo un coeficiente de determinación $R^2 = 0.0709$ y una correlación $r = 0.2663$, mientras que el modelo polinomial inverso, representado como $Y = a / (b \cdot X^2) + c \cdot X$, presentó

resultados muy similares ($R^2 = 0.0707$, $r = 0.2660$). Ambos modelos muestran niveles bajos de ajuste y correlación, aunque las métricas reflejan una ligera mejora respecto a otras relaciones analizadas. Estos valores indican una correlación débil pero positiva, lo que sugiere que el avance porcentual del proyecto tiene una influencia limitada sobre su estado general de salud.

Visualmente, las gráficas exhiben una alta dispersión de puntos y la ausencia de agrupamientos definidos o tendencias claras, lo que refuerza la conclusión de que el progreso porcentual no implica necesariamente una mejora proporcional en la salud del proyecto. Esto puede deberse a la influencia de otros factores, como la calidad de la ejecución, la gestión de riesgos o el cumplimiento de hitos clave.

Función R^2 :0.0709

Correlación (r):-0.1784

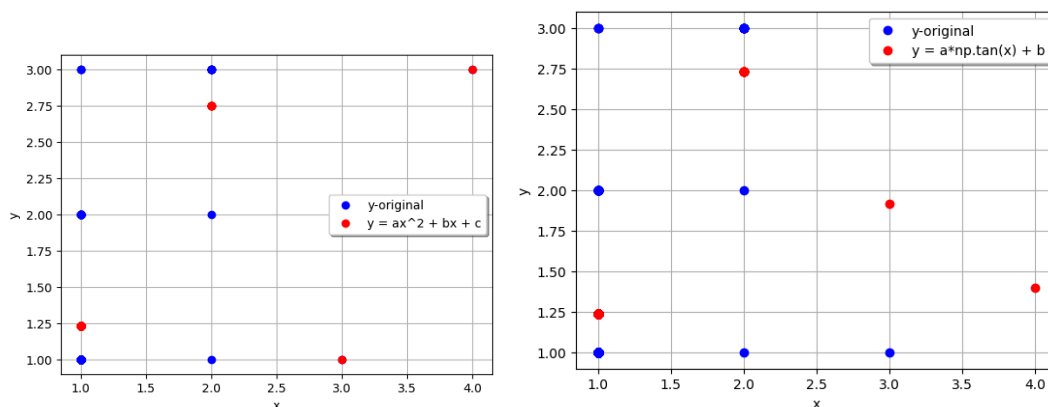
Correlación Final (r):0.2663

Project Health

State - Project Health

State - Independiente

Project Health - Dependiente



Análisis entre Project Health y State

Se aplicaron dos modelos de regresión no lineal —cuadrático y de cociente entre polinomios— con el objetivo de analizar la relación entre el estado general del proyecto (Project Health) y su estado operativo (State).

El modelo cuadrático, expresado por la ecuación $Y = a \cdot X^2 + b \cdot X + c$, presentó un coeficiente de determinación $R^2 = 0.3045$ y una correlación $r = 0.5518$, mientras que el modelo de cociente entre polinomios, definido como $Y = (a \cdot X^2 + b) / (c \cdot X^2 + d)$, obtuvo resultados muy similares ($R^2 = 0.3154$, $r = 0.5616$).

Estos valores indican una correlación moderada y positiva, con un comportamiento semejante en ambos modelos. Sin embargo, la pendiente observada es prácticamente nula, lo que sugiere que no existe una tendencia clara entre el estado operativo y la salud general del proyecto. Esto implica que las variaciones en Project Health pueden deberse a factores externos como la gestión de recursos, la planificación o la mitigación de riesgos, más que al estado operativo en sí.

Visualmente, las gráficas muestran una alta dispersión de puntos y la ausencia de agrupamientos definidos, lo que refuerza la idea de que la relación entre ambas variables es no lineal y débil.

Aun así, el modelo de cociente entre polinomios presenta un ligero mejor ajuste global ($R^2 = 0.3154$), por lo que se considera el más adecuado dentro del bajo nivel de correlación observado.

Función R^2 :0.3154

Correlación (r):0.4938

Correlación Final (r):0.5616

Conclusiones

Tabla 2. Coeficientes de correlación (r) y determinación (R^2) obtenidos en los modelos de regresión no lineal

Variables	Función aplicada	Correlación (r)	R^2 (Coef. de determinación)
Project Type vs BG	Cuadrática	0.3876	0.1502
Project Type vs BG	Exponencial	0.0000	0.0000
Geographical scope vs Project manager	Cuadrática	0.2068	0.0428

Geographical scope vs Project manager	Cociente Polinomios	0.1563	0.0244
Project manager vs Project organization	Cuadrática	0.5205	0.2709
Project manager vs Project organization	Logarítmica	0.5290	0.2798
State vs Project Health	Senoidal	0.6688	0.4473
State vs Project Health	Valor Absoluto	0.4938	0.2438
Project size vs On-hold	Cuadrática	0.1949	0.0380
Project size vs On-hold	Valor Absoluto	0.1375	0.0189
Project organization vs Project manager	Cuadrática	0.5101	0.2602
Project organization vs Project manager	Valor Absoluto	0.5047	0.2548
BG vs Project Type	Cociente Polinomios	0.4613	0.2128
BG vs Project Type	Logarítmica	0.4516	0.2039

On-hold vs Project Health	Cuadrática	0.3433	0.1179
On-hold vs Project Health	Valor Absoluto	0.3344	0.1118
Percent complete vs Project Health	Cuadrática	0.2663	0.0709
Percent complete vs Project Health	Polinomial Inversa	0.2660	0.0708
Project Health vs State	Cuadrática	0.5519	0.3046
Project Health vs State	Cociente Polinomios	0.5616	0.3154