



Tecnológico de Monterrey

Actividad 3.2 Regresión No Lineal

Lucero Jannete López García A01736938

Pilar Méndez Briones A01736843

Rubén Flores A01737098

Braulio González

**Analítica de datos y herramientas de inteligencia artificial
I (Gpo 101)**

Fecha

08/04/2025

Analizar las correlaciones lineales simples (“Heatmap”) que tiene la columna “Objetivo” (Libre a su elección) con respecto a las demás variables numéricas, proponer un modelo No lineal que supere la mejor correlación lineal simple encontrada para esta variable. Argumentar los hallazgos obtenidos en este punto.

Nosotros tomamos como variable objeto BOTÓN CORRECTO, y al compararlo con sus variables independientes, encontramos que la que tiene mayor correlación es número de interacción, sin embargo, realizamos la correlación con modelos NO lineales específicamente la función cuadrática con todas las variables, encontrando que esta función mejora las correlaciones entre 1 a 0.3 o en caso de autopush, tiempo de sesión y tiempo de lección no se encuentra relación en ninguna de las variables.

| Atributo | Correlación Modelo 1 simple regresión lineal | Correlación Modelo 2 cuadrática |
|-----------------------|---|---------------------------------|
| Administrador | 0.151109 | 0.18 |
| Usuario | 0.173722 | 0.21 |
| mini juego | 0.232161 | 0.29 |
| color presionado | 0.456837 | 0.59 |
| dificultad | 0.021313 | 0.03 |
| juego | 0.055934 | 0.05 |
| tiempo de interacción | 0.279734 | 0.3567 |
| número de interacción | 0.467015 | 0.6033 |
| auto push | NaN | NaN |
| tiempo de sesión | NaN | NaN |
| tiempo de lección | NaN | NaN |

Al analizar los datos de correlación entre los distintos atributos y los modelos de regresión (modelo lineal y modelo cuadrático), identificamos varios puntos clave que consideramos importantes para destacar.

En primer lugar, observamos que el número de interacciones fue el atributo con la correlación más alta en ambos modelos. En el modelo lineal obtuvo un valor de 0.467, y en el modelo cuadrático aumentó a 0.6033. Esto indica que la cantidad de veces que los usuarios interactúan con el sistema influye directamente en el resultado del modelo, siendo un factor clave para entender su comportamiento.

En segundo lugar, el atributo “color presionado” también presentó una correlación elevada: 0.4568 en el modelo lineal y 0.59 en el cuadrático. Esto sugiere que las elecciones visuales

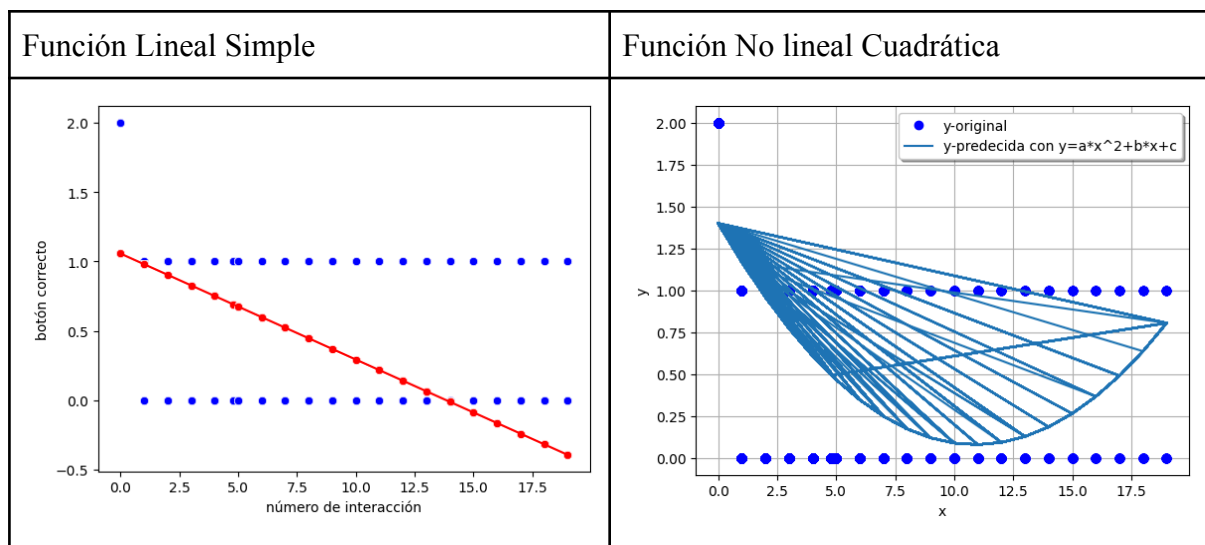
realizadas por los usuarios, como los colores seleccionados durante la interacción, tienen un impacto significativo en los resultados del análisis.

Además, notamos que el modelo cuadrático mejora la correlación en todos los atributos respecto al modelo lineal. Por ejemplo, el atributo “usuario” pasó de 0.1737 a 0.21, “mini juego” de 0.2321 a 0.29 y “tiempo de interacción” de 0.2797 a 0.3567. Esto refleja que la relación entre los atributos y la variable dependiente no es completamente lineal, por lo que el modelo cuadrático logra capturar de mejor forma las variaciones en los datos.

Por otro lado, identificamos algunos atributos que no aportan información relevante al modelo o carecen de datos suficientes, como “auto push”, “tiempo de sesión” y “tiempo de lección”, los cuales presentan valores “NaN”. Asimismo, atributos como “dificultad” y “juego” tienen una correlación muy baja (0.02–0.05), lo que nos lleva a considerar que su influencia en el modelo es mínima.

Sí bien hicimos de todas, es importante solo poder comparar la que cuente con mayor correlación, como la actividad marca para observar el comportamiento de cada modelo con funciones lineales contra las funciones no lineales como lo podemos observar a continuación.

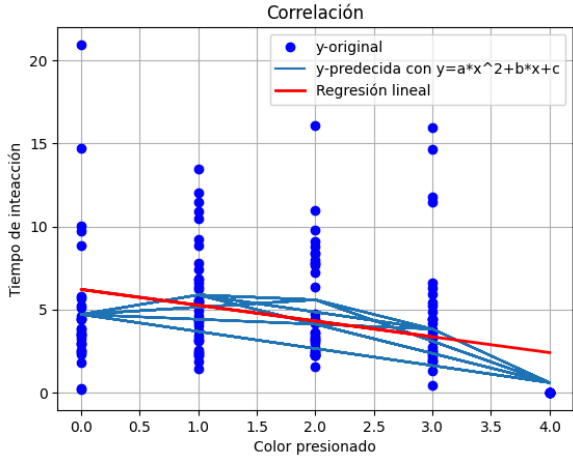
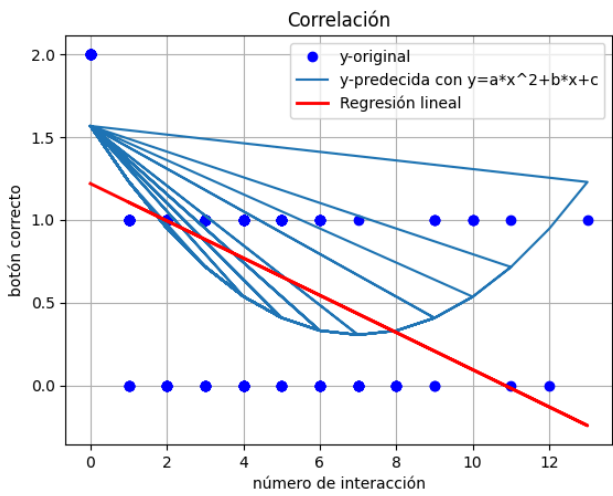
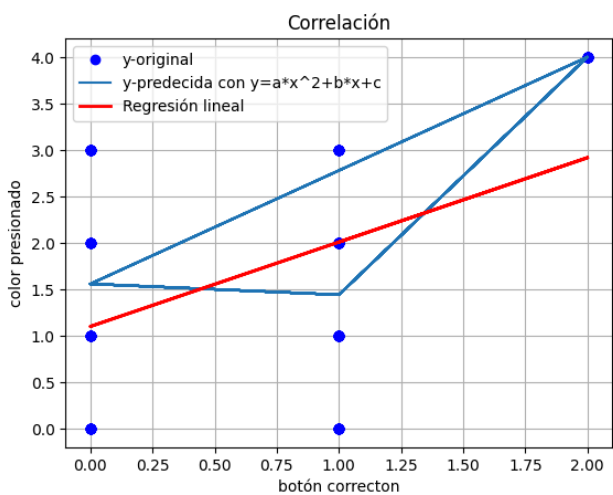
Botón correcto vs Número de interacción

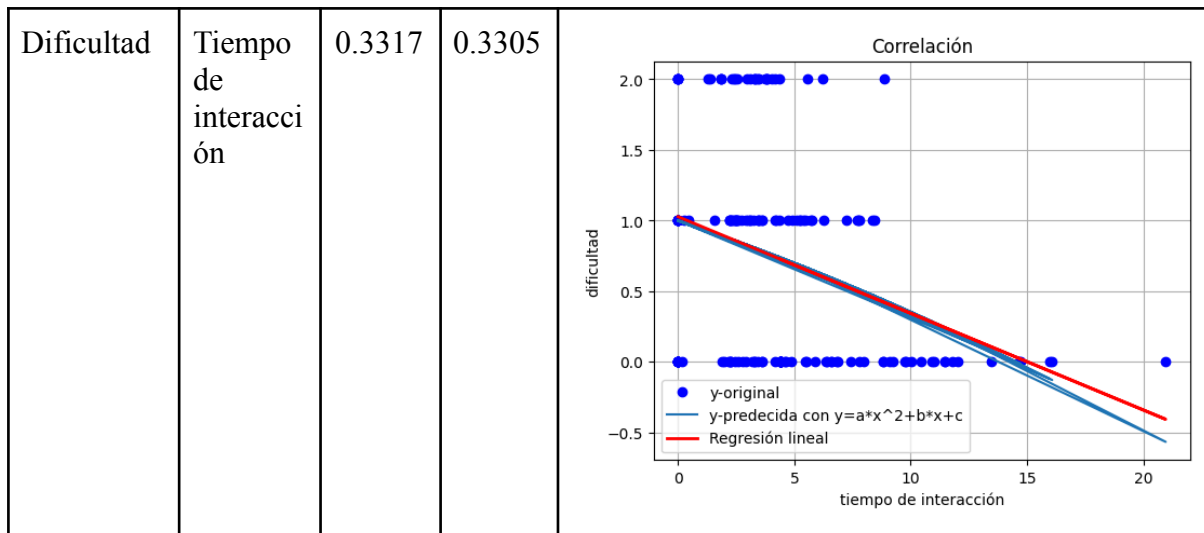


Como podemos observar la función cuadrática abarca más áreas de los datos causando que tenga mejor exactitud y relación entre las variables

Por usuario

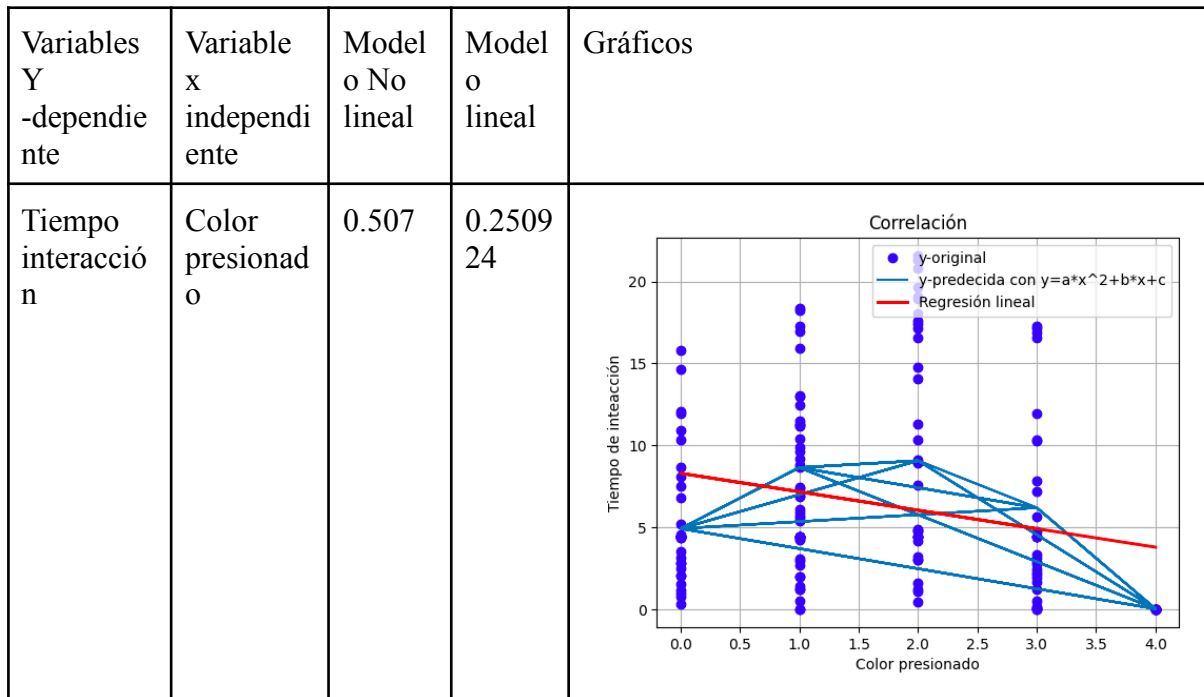
Valentin

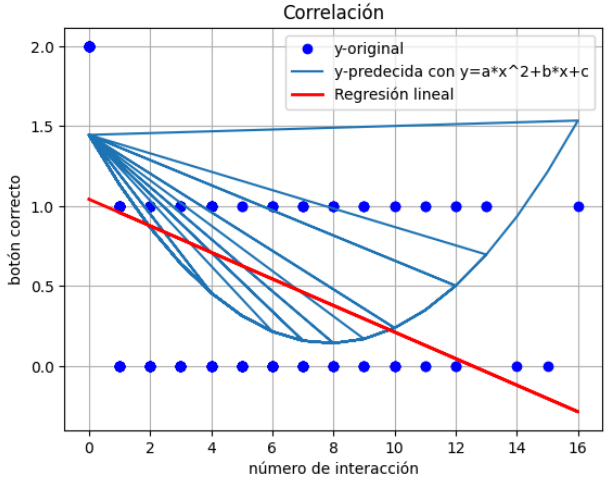
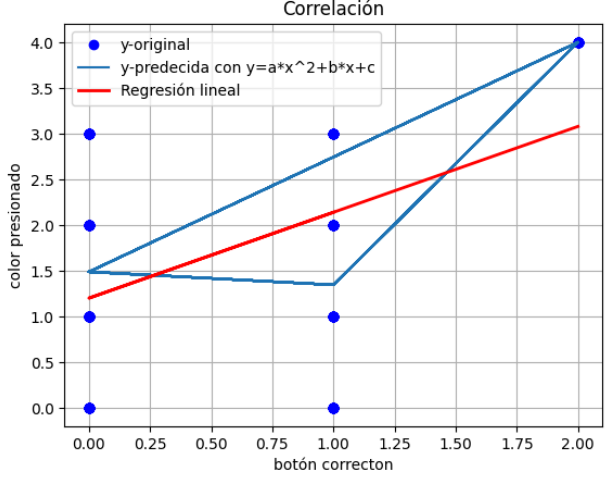
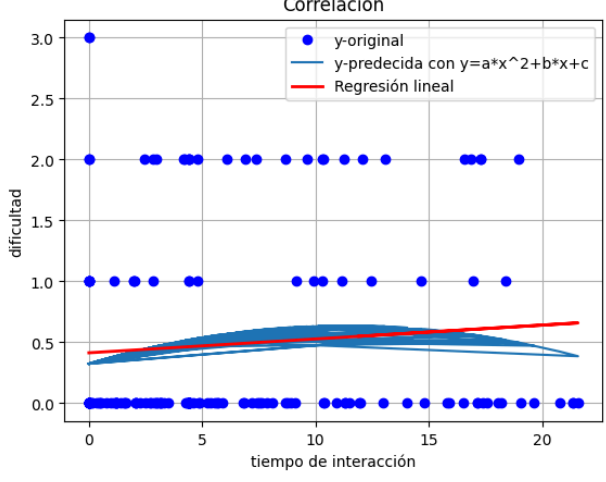
| Variables Y-dependiente | Variable x indep | Modelo No lineal | Modelo lineal | Gráficos |
|-------------------------|-----------------------|------------------|---------------|--|
| Tiempo interacción | Color presionado | 0.46 | 0.3366 |  |
| Botón correcto | Número de interacción | 0.641 | 0.4917 |  |
| Color presionado | Botón correcto | 0.662 | 0.45 |  |



- El modelo no lineal mejora significativamente los resultados en todas las variables.
- La correlación más alta es entre color presionado y botón correcto (0.662), lo que indica una fuerte relación entre la acción del usuario y la precisión.
- El modelo cuadrático ofrece un mejor ajuste para casi todos los pares de variables, especialmente botón correcto vs número de interacción.

Yael david

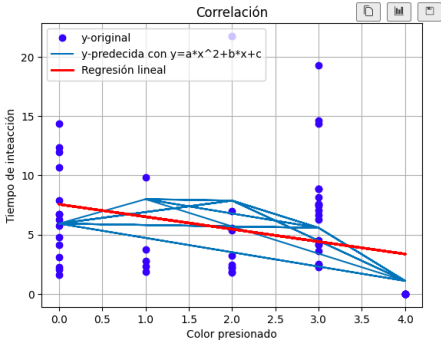
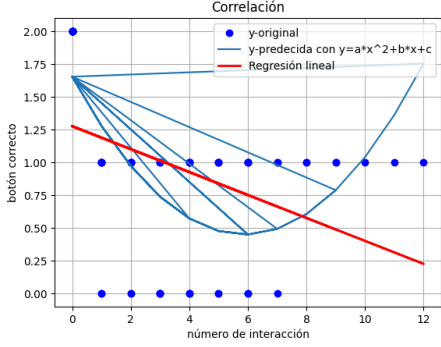
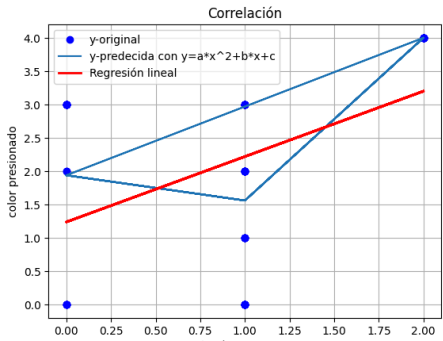


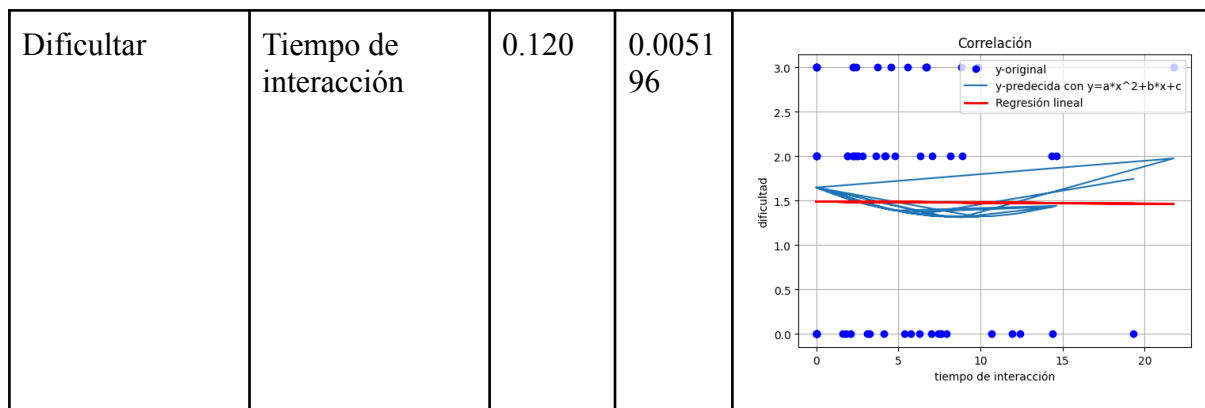
| | | | | |
|------------------|-----------------------|--------|-------|--|
| Botón correcto | Número de interacción | 0.6323 | 0.425 |  |
| Color presionado | Botón correcto | 0.689 | 0.50 |  |
| Dificultar | Tiempo de interacción | 0.134 | 0.08 |  |

- Presenta la correlación más alta en todo el análisis grupal (0.689) entre color presionado y botón correcto en el modelo no lineal.
- La relación entre dificultad y tiempo de interacción es muy baja, lo que indica que este atributo no tiene gran peso en su caso.

- También se observa una mejora considerable al usar el modelo no lineal, sobre todo en “tiempo de interacción”.

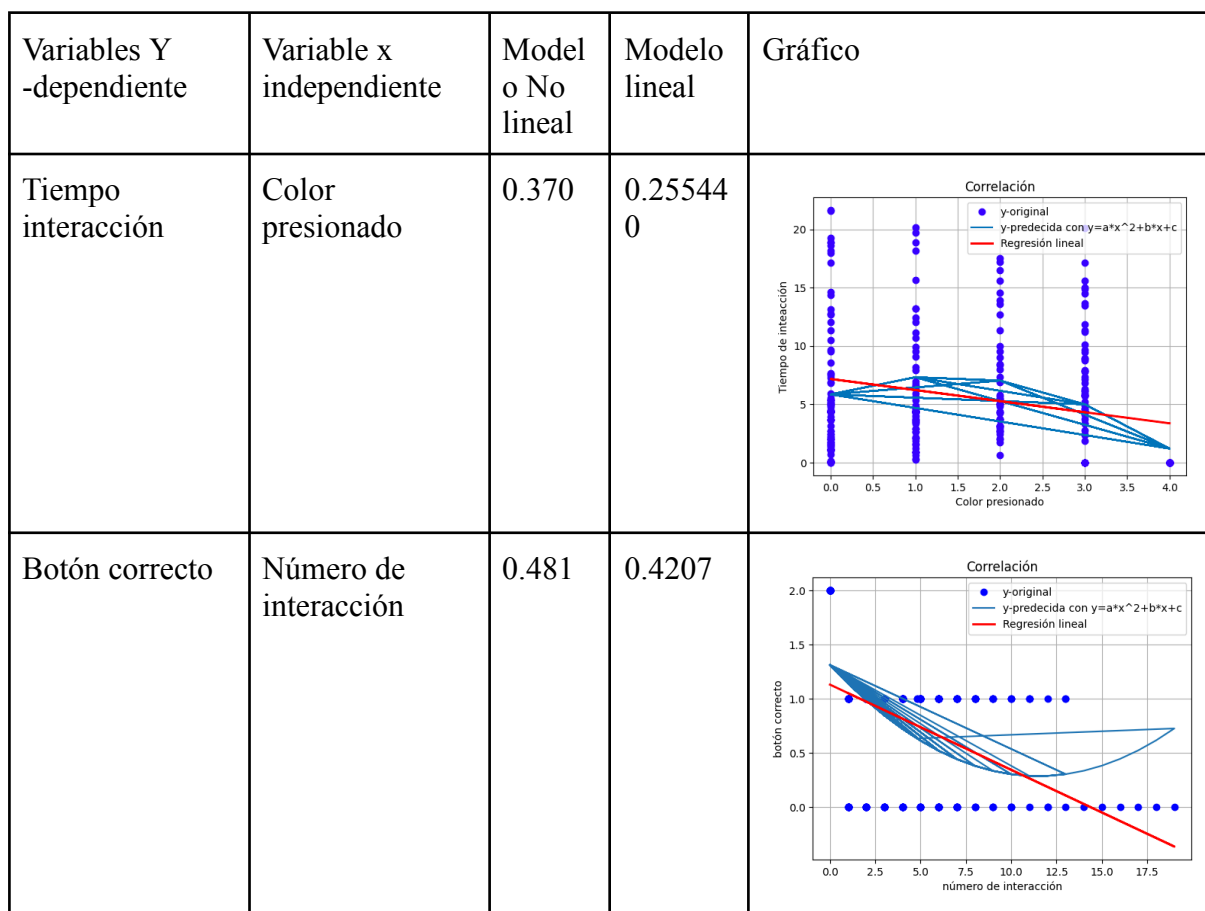
Yeremy Jazmin

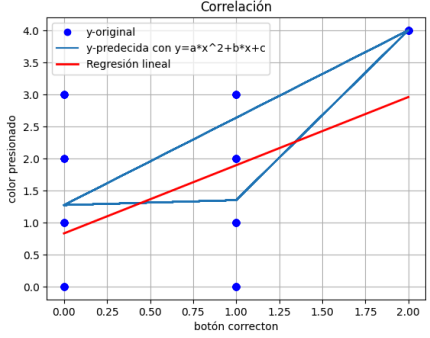
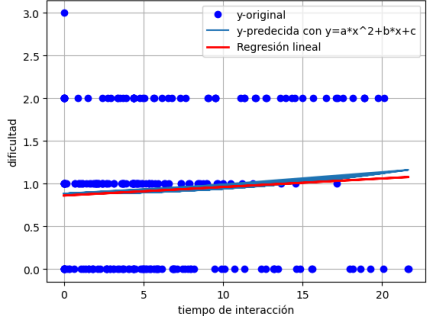
| Variables Y -dependiente | Variable x independiente | Model o No lineal | Model o lineal | Gráficos |
|-----------------------------|-----------------------------|-------------------------|----------------------|--|
| Tiempo interacción | Color presionado | 0.455 | 0.29 |  |
| Botón correcto | Número de interacción | 0.673 | 0.38 |  |
| Color presionado | Botón correcto | 0.662 | 0.4574 |  |



- Tiene valores similares a los de Yael y Valentín.
- Su mayor correlación está en botón correcto vs número de interacción (0.673) en el modelo no lineal.
- La relación dificultad vs tiempo de interacción es la más baja del grupo, especialmente en el modelo lineal (0.0052), prácticamente nula.

Ramiro



| | | | | |
|------------------|-----------------------|--------|--------|--|
| Color presionado | Botón correcto | 0.658 | 0.4956 |  |
| Difícultar | Tiempo de interacción | 0.0704 | 0.065 |  |

- Aunque sus correlaciones son un poco más bajas que las de los otros tres usuarios anteriores, mantiene la misma tendencia: el modelo no lineal mejora todos los casos.
- La mayor correlación es también en color presionado vs botón correcto (0.658).
- La relación entre dificultad y tiempo de interacción es muy baja, pero un poco más estable que en el caso de Jeremy.

Irving

| Variables Y -dependiente | Variable x independiente | Modelo No lineal | Modelo lineal |
|--------------------------|--------------------------|------------------|---------------|
| Tiempo interacción | Color presionado | Error | NAN |
| Botón correcto | Número de interacción | Error | NAN |
| Color presionado | Botón correcto | Error | NAN |
| Difícultar | Tiempo de interacción | Error | NAN |

- No se pudo obtener ningún dato de correlación para este usuario; todos los valores están marcados como Error o NaN, lo que impide su análisis.

En todos los casos (excepto Irving), el modelo no lineal supera al modelo lineal, confirmando que las relaciones entre variables no son del todo lineales y se ajustan mejor a una función cuadrática.

- Las variables más fuertes en todos los usuarios fueron:
- Color presionado vs botón correcto (máxima correlación general).
- Botón correcto vs número de interacción (en todos los casos > 0.48 en modelo no lineal).
- Las variables menos relevantes fueron:
- Dificultad vs tiempo de interacción, con correlaciones consistentemente bajas, incluso en el modelo no lineal.
- Esto sugiere que la dificultad percibida no depende directamente del tiempo que el usuario tarda.

Conclusión

Encontramos que el modelo cuadrático o no lineal representa de manera más precisa las relaciones entre las variables, ya que en todos los casos analizados mostró un incremento significativo en las correlaciones respecto al modelo lineal. Esto indica que las relaciones entre los datos no siguen un patrón estrictamente lineal, sino que presentan una curvatura que el modelo cuadrático logra capturar. Por ejemplo, pequeñas variaciones en el número de interacciones o en el color presionado pueden tener un impacto mayor o menor dependiendo del contexto, algo que el modelo lineal no logra reflejar con la misma precisión.

Finalmente, esta información nos resulta útil para construir modelos predictivos más robustos y tomar decisiones informadas sobre qué variables priorizar en futuras mejoras del sistema. También resaltamos la necesidad de revisar la base de datos de Irving para mantener la coherencia y confiabilidad del análisis grupal.