Universidad
Francisco de Vitoria
UFV Madrid

2ª Parte del Trabajo de Fin de Grado

Análisis de los Datos

Trabajo Fin de Grado
Grado en Análisis de Negocios – Business Analytics
Febrero de 2024

Autor:
Ignacio López de Carrizosa Grosso

Tutor:
Prof. Dra. Ana Lazcano de Rojas

Facultad de Facultad de Derecho, Empresa y Gobierno
Universidad Francisco de Vitoria

RESUMEN

Este Trabajo de Fin de Grado (TFG) se originó con un marcado interés en analizar los efectos de la pandemia de COVID-19 sobre el entramado empresarial en España, pero evolucionó hacia una investigación más exhaustiva sobre cómo diversas variables macroeconómicas afectan al sector empresarial del país.

La primera fase del proyecto, denominada "Ingeniería del Dato", se dedicó a la meticulosa tarea de Extracción, Transformación y Carga (ETL) de datos relevantes. Este proceso fue esencial para asegurar la calidad y relevancia de los datos para análisis posteriores. La recopilación de datos se llevó a cabo con la ayuda de fuentes confiables como el Instituto Nacional de Estadística (INE) y epdata, proporcionando una base sólida para las etapas subsiguientes del estudio.

Posteriormente, el proyecto transitó hacia la segunda fase, titulada "Análisis del Dato", donde el enfoque se amplió para incluir un análisis más general de la influencia de las variables macroeconómicas sobre el sistema empresarial. Este cambio en el enfoque del TFG refleja una adaptación y evolución del objetivo inicial, considerando no solo el impacto directo de la pandemia, sino también la forma en que el entorno económico y las políticas gubernamentales configuran el panorama empresarial en España.

Este enfoque más amplio permitió una exploración detallada de las dinámicas empresariales, revelando interacciones complejas entre la creación de empresas, el capital desembolsado y variables clave como el ICEA o el IPC. Utilizando técnicas de análisis de datos, se identificaron patrones y correlaciones significativas que ofrecen perspectivas valiosas sobre la estabilidad y la capacidad de adaptación del tejido empresarial en España.

El resumen narrativo compila observaciones y hallazgos de ambos enfoques, subrayando la necesidad de considerar un amplio rango de factores al examinar los impactos en el sistema empresarial. La combinación de un enfoque inicial centrado específicamente en la pandemia, seguido de un análisis más expansivo de las variables macroeconómicas, proporciona una visión más completa y matizada de las dinámicas empresariales en España.

El proyecto concluye con una serie de recomendaciones basadas en el análisis efectuado. Este TFG no solo enriquece el cuerpo académico existente sobre las disoluciones y creaciones empresariales en España, sino que también ofrece herramientas analíticas que pueden ser útiles en investigaciones futuras y en la práctica empresarial. Este estudio destaca la importancia de abordajes integrales y bien fundamentados para fomentar un entorno empresarial estable y propicio al crecimiento sostenido en el cambiante contexto económico y social de España.

Índice del Documento

SEGUNDA PARTE: ANÁLISIS DEL DATO

8. Introducción al análisis del dato

El análisis de datos se ha convertido en una herramienta fundamental en el ámbito de la investigación y la toma de decisiones en diversas disciplinas, permitiendo extraer conocimientos valiosos a partir de grandes volúmenes de información.

Este documento se enfoca en explorar y aplicar técnicas avanzadas de análisis de datos para comprender mejor las dinámicas y factores que influyen en las disoluciones empresariales en España, un tema de gran relevancia económica y social.

A lo largo de este trabajo, se seleccionarán y prepararán cuidadosamente los datos para su análisis, asegurando que la información sea precisa y esté lista para ser examinada a través de diversos modelos analíticos. Estos modelos, desarrollados y aplicados meticulosamente, buscarán identificar patrones, correlaciones y posibles causas detrás de las disoluciones empresariales, utilizando para ello un enfoque multidimensional que incluye variables económicas, sociales y de mercado. Se evaluará la adecuación de los modelos empleados mediante medidas estadísticas que permitan verificar su fiabilidad y precisión. Además, se hará uso de técnicas de visualización de datos para presentar los resultados de manera clara y comprensible, facilitando así la interpretación de los hallazgos.

Finalmente, este análisis culminará en la elaboración de conclusiones y recomendaciones basadas en los resultados obtenidos, proporcionando insights valiosos. Este trabajo además de tratar aportar al conocimiento académico sobre las disoluciones empresariales en España intentará ofrecer herramientas analíticas que puedan ser aplicadas en futuras investigaciones y en la práctica empresarial para fomentar un entorno económico más estable y resiliente.

9. Selección y Preparación de Datos para el Análisis

La selección y preparación de datos constituyen etapas cruciales en el proceso de análisis de datos, especialmente cuando se abordan cuestiones complejas como las situaciones que influyen en las disoluciones empresariales. Este trabajo se ha fundamentado en el análisis exhaustivo de tres bases de datos principales, cada una de ellas derivada y refinada a partir de conjuntos de datos más amplios, con el objetivo de explorar distintas facetas del fenómeno en estudio.

La primera base de datos, inicialmente denominada "dfbbdd1" y posteriormente segmentada en "numsoc" y "capdes", se centró en recopilar información relativa al número de sociedades creadas por tipo de sociedad, año y comunidad autónoma, así como el capital desembolsado por estas sociedades, complementado con datos demográficos por comunidad autónoma. Esta división permitió abordar dos análisis de regresión lineal simple, orientados a evaluar el impacto de la población en la creación de empresas y en el capital desembolsado, respectivamente, proporcionando una visión detallada de cómo la demografía puede influir en el tejido empresarial.

En el desarrollo de la investigación para este Trabajo de Fin de Grado, se configuraron dos bases de datos fundamentales para el análisis de las dinámicas empresariales en España. La primera, denominada "disolución", compila el número de empresas disueltas por tipo de disolución en cada comunidad autónoma desde el año 2012 hasta el 2022. Complementariamente, se creó una segunda base de datos, llamada "macrospain", que recoge indicadores económicos esenciales como el Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) por comunidad autónoma y año. El objetivo de vincular estos conjuntos de datos era llevar a cabo una regresión lineal múltiple para investigar cómo las condiciones económicas impactan en la tasa de disolución empresarial.

De manera paralela y con la misma metodología, se desarrolló otra base de datos, "regmul2", destinada a registrar el número de empresas constituidas en cada comunidad autónoma durante el mismo periodo. Esta base de datos se orienta a explorar la relación entre los mismos indicadores económicos y la tasa de creación de empresas, permitiendo un análisis comparativo entre los factores que influyen tanto en la constitución como en la disolución de empresas dentro del territorio español. Esta dualidad de bases de datos enriquece significativamente el estudio, permitiendo una comprensión más completa de las fuerzas económicas que moldean el panorama empresarial del país.

Para profundizar en el análisis de las dinámicas empresariales en España, se creó una tercera base de datos denominada "dismac", destinada a ampliar el espectro de variables macroeconómicas analizadas en relación con las disoluciones empresariales. Esta base de

datos incluye indicadores tales como la deuda pública, el déficit público, el gasto público, los ingresos fiscales, el turismo y las reservas nacionales, junto con el Índice de Confianza Empresarial Armonizado (ICEA), para cada comunidad autónoma y año. Este enfoque multidimensional ofrece una visión más rica y detallada de cómo diversos factores económicos y sociales pueden influir en la estabilidad y continuidad de las empresas en España.

De manera similar, se creó una base de datos complementaria llamada "consmac", que aplica la misma metodología pero se centra en las constituciones empresariales. "consmac" también integra las mismas variables macroeconómicas y busca explorar cómo estos factores influyen en la creación de nuevas empresas, permitiendo un análisis comparativo entre los impulsores de las constituciones y disoluciones empresariales. Este enfoque paralelo facilita una comprensión integral de los efectos de las políticas económicas y las condiciones de mercado en el tejido empresarial del país.

Por último, para complementar y enriquecer aún más el análisis sobre las dinámicas empresariales en España, se han incorporado dos nuevas bases de datos: "consext" y "disext", correspondientes a las constituciones y disoluciones empresariales, respectivamente. Estas bases de datos han sido diseñadas para explorar la relación entre un amplio espectro de variables y las dinámicas de constitución y disolución de empresas. Incluyen variables tanto económicas como sociales, ampliando considerablemente el rango de factores considerados en el análisis. Entre estas variables se encuentran la cotización del IBEX, del EUR y del NASDAQ, que reflejan el comportamiento del mercado financiero; indicadores macroeconómicos como el déficit público, el gasto público y los ingresos fiscales; y variables demográficas como nacimientos, matrimonios, defunciones y población total.

La inclusión de estas variables busca proporcionar una visión más holística y detallada de los factores que pueden influir en la creación y cierre de empresas en España. Se parte del supuesto de que a mayor número de variables analizadas, mayor será la capacidad del modelo para explicar las fluctuaciones en las tasas de constituciones y disoluciones empresariales. Con "consext" y "disext", se pretende no solo identificar las tendencias y patrones más evidentes, sino también descubrir conexiones menos obvias que puedan surgir de la interacción entre el entorno económico, el mercado financiero y la dinámica social, ofreciendo así una comprensión más profunda y matizada de los desafíos y oportunidades dentro del ecosistema empresarial español.

Cada una de estas bases de datos fue sometida a un riguroso proceso de selección y preparación, que incluyó la limpieza de datos, la gestión de valores faltantes, la transformación de variables y la verificación de la calidad de los datos. Este proceso aseguró que la información utilizada en los análisis fuera de la más alta calidad y relevancia, permitiendo así

obtener resultados confiables. La meticulosa preparación de los datos subraya la importancia de una base sólida para cualquier análisis de datos, especialmente cuando se abordan cuestiones de complejidad y relevancia como las que conciernen a las disoluciones empresariales en el contexto económico y social de España.

10. Modelos Analíticos: Desarrollo y Aplicación (Regresiones Lineales)

10.1. Modelo Analítico Supervisado (Regresiones Lineales Simples)

Explicación del Modelo:

En el análisis del comportamiento empresarial y económico, la regresión lineal simple emerge como una herramienta analítica fundamental, especialmente cuando el objetivo es explorar la relación entre dos variables específicas. Este modelo supervisado se seleccionó con el propósito de investigar cómo la variable independiente, en este caso, la población de una comunidad autónoma puede influir en la variable dependiente, que para la primera regresión se define como el número de sociedades creadas y para la segunda como el capital desembolsado en miles de euros.

La regresión lineal simple es un método estadístico fundamental utilizado para modelar y analizar las relaciones entre dos variables cuantitativas: una variable independiente (o explicativa) y una variable dependiente (o respuesta). Este modelo asume que existe una relación lineal entre estas variables, la cual puede ser descrita mediante una ecuación de la forma:

$$\boldsymbol{y = \beta_0 + \beta_1 x + \epsilon}$$

Donde:

- $\boldsymbol{y}$ representa la variable dependiente.
- $\boldsymbol{x}$ representa la variable independiente.
- $\boldsymbol{\beta_0}$ es el término de intercepción, que indica el valor de y cuando x es 0.
- $\boldsymbol{\beta_1}$ es el coeficiente de la pendiente, que indica el cambio en y por cada unidad de cambio en x.
- $\boldsymbol{\epsilon}$ es el término de error, que representa la variación en y que no puede ser explicada por la relación lineal con x.

El método de mínimos cuadrados es utilizado para estimar los parámetros $\boldsymbol{\beta 0}$ y $\boldsymbol{\beta 1}$ de la regresión. Este método busca minimizar la suma de los cuadrados de las diferencias (residuos) entre los valores observados de "$\boldsymbol{y}$" y los valores predichos por el modelo. Matemáticamente, se busca minimizar la función:

$$\boldsymbol{S(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2}$$

Donde $\boldsymbol{yi}$ y $\boldsymbol{xi}$ son los valores observados de la variable dependiente e independiente, respectivamente, y n es el número de observaciones. La solución a este problema de optimización nos da los estimadores de mínimos cuadrados para $\boldsymbol{\beta 0}$ y $\boldsymbol{\beta 1}$, que son:

$$\boldsymbol{\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\,(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

$$\boldsymbol{\widehat{\beta}_0 = \overline{y} - \beta_1 \overline{x}}$$

Donde $\boldsymbol{\overline{x}}$ e $\boldsymbol{\overline{y}}$ son las medidas de las variables independiente y dependiente, respectivamente.

El $\boldsymbol{\beta 1}$ coeficiente indica la pendiente de la línea de regresión y representa el cambio esperado en $\boldsymbol{y}$ por cada unidad de cambio de $\boldsymbol{x}$. Un valor positivo de $\boldsymbol{\beta_1}$ indica una relación directa entre $\boldsymbol{x}$ e $\boldsymbol{y}$, mientras que un valor negativo indica una relación inversa.

El término de intercepción $\boldsymbol{\beta 0}$ indica el valor esperado de $\boldsymbol{y}$ cunado $\boldsymbol{x}$ es $\boldsymbol{0}$. Este valor puede no tener siempre una interpretación práctica, especialmente si $\boldsymbol{x = 0}$ no se encuentra dentro del rango de los datos observados.

Para que las estimaciones obtenidas mediante la regresión lineal simple sean válidas, se deben cumplir ciertos supuestos:

1) $\boldsymbol{Linealidad}$: La relación entre $\boldsymbol{x}$ e $\boldsymbol{y}$ debe ser lineal.
2) $\boldsymbol{Independencia}$: Los residuos $\boldsymbol{\varepsilon}$ deben ser independientes entre sí.
3) $\boldsymbol{Homoscedasticidad}$: La varianza de los residuos debe ser constante a lo largo de todas las observaciones.
4) $\boldsymbol{Normalidad}$: Los residuos deben seguir una distribución normal.

Proceso de Selección de Variables:

La selección de variables para la regresión lineal simple se centró en identificar la variable independiente (X) y la variable dependiente (Y) que mejor representaran la relación que se deseaba explorar. Dada la naturaleza del análisis, se identificó la población de las comunidades autónomas como la variable independiente, considerando su potencial impacto en el atractivo para la creación de empresas y la capacidad económica de estas. Por otro lado, se eligió el número de sociedades creadas como la variable dependiente de la primera regresión y al capital desembolsado en miles de euros por las sociedades para la segunda regresión. Para llevar a cabo el análisis, se dividió la base de datos original en dos subconjuntos: uno enfocado en el número de sociedades y otro en el capital desembolsado, permitiendo así un estudio detallado y específico de cada aspecto. Este proceso de selección de variables fue crucial para asegurar que el modelo pudiera capturar de manera efectiva la relación entre la demografía de las comunidades autónomas y la actividad económica empresarial, facilitando la interpretación de los resultados y la extracción de conclusiones relevantes.

Desarrollo de modelos:

Regresión 1: Número de Sociedades

- *Análisis de la Existencia de Relación Lineal:* Para explorar la relación entre la población de las comunidades autónomas y el número de sociedades creadas, se realizaron análisis gráficos preliminares. Se emplearon gráficos de dispersión con líneas de tendencia suavizadas para visualizar la distribución de los datos y detectar patrones de correlación visualmente. Estos gráficos permitieron una primera aproximación a la dinámica entre las variables, sugiriendo una relación que, a primera vista, podría no ser lineal o ser muy débil, dada la dispersión de los puntos y la suavidad de la línea de tendencia. Para complementar el análisis gráfico, se calculó la correlación entre 'Población' y 'Número de Sociedades' utilizando la función corr() de pandas, seguido de un análisis más formal mediante el coeficiente de correlación de Pearson y su p-valor asociado. Los resultados indicaron una correlación de -0.008 con un p-valor de 0.761, lo que sugiere que, estadísticamente, no existe una relación lineal significativa entre la población y el número de sociedades creadas.
- *Análisis de Ajuste a una Distribución Normal:* El ajuste de las variables a una distribución normal es crucial para la aplicación de ciertas técnicas estadísticas.

Se utilizó la visualización mediante gráficos de densidad y se realizaron pruebas de normalidad, incluyendo Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2. Los gráficos de densidad revelaron una distribución asimétrica para ambas variables, confirmada por los valores de asimetría (skewness) significativamente diferentes de cero. Además, las pruebas de normalidad arrojaron p-valores extremadamente bajos para ambas variables, indicando un rechazo de la hipótesis nula de normalidad.

- *Construcción del Modelo:* Para la construcción del modelo de regresión lineal simple, se prepararon las variables seleccionadas, añadiendo una columna de unos para el intercepto. A pesar de la aparente falta de una relación lineal significativa y la no normalidad de las distribuciones, se procedió a ajustar el modelo para explorar la relación entre las variables de interés.
  El modelo ajustado mostró un R-cuadrado cercano a cero, indicando que el modelo no explica prácticamente ninguna variabilidad en el número de sociedades en función de la población. Los coeficientes de regresión y sus intervalos de confianza reflejaron la falta de significancia estadística de la población como predictor del número de sociedades.

Regresión 2: Capital Desembolsado

- *Análisis de la Existencia de Relación Lineal:* Para investigar la relación entre la población de las comunidades autónomas y el capital desembolsado en la creación de sociedades, se emplearon gráficos de dispersión complementados con líneas de tendencia suavizadas. Estos gráficos facilitaron una visualización preliminar de la relación entre las variables, sugiriendo la necesidad de un análisis más detallado.
  La correlación entre 'Población' y 'Capital' se calculó utilizando la función corr() de pandas, y se complementó con el coeficiente de correlación de Pearson y su p-valor asociado. Los resultados mostraron una correlación de 0.20 con un p-valor significativamente bajo (4.639e-14), lo que indica una relación lineal positiva estadísticamente significativa entre la población y el capital desembolsado, aunque la fuerza de esta relación es moderada.
- *Análisis de Ajuste a una Distribución Normal:* El análisis de la distribución de las variables mediante gráficos de densidad y pruebas de normalidad reveló una distribución asimétrica para ambas variables, lo que se reflejó en los valores de asimetría significativamente altos. Las pruebas de Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2 confirmaron la no normalidad de las

distribuciones, con p-valores que indican un rechazo fuerte de la hipótesis nula de normalidad.

- *Construcción del Modelo:* A pesar de la moderada correlación positiva entre la población y el capital desembolsado y la no normalidad de las distribuciones, se procedió a ajustar un modelo de regresión lineal simple. Se prepararon las variables seleccionadas, incluyendo una columna de unos para el intercepto, y se ajustó el modelo para explorar la relación entre la población y el capital desembolsado.

  El modelo ajustado reveló un R-cuadrado de 0.043, indicando que un 4.3% de la variabilidad en el capital desembolsado puede explicarse por la población. Aunque esta proporción es baja, el coeficiente para la población fue estadísticamente significativo, lo que sugiere que existe una relación lineal positiva entre la población y el capital desembolsado. Este análisis resalta que aunque la relación entre la población y el capital desembolsado es estadísticamente significativa y la fuerza de esta relación es moderada, la no normalidad de las variables necesita un análisis más profundo.

Evaluación del Modelo con Medidas de Error/Precisión Específicas:

Regresión 1: Número de Sociedades

- *Calidad del Modelo:* El R-cuadrado obtenido en el modelo es 0.000, indicando que la variabilidad explicada por el modelo es prácticamente nula. Este valor sugiere que la población, como variable independiente, no proporciona una base sólida para predecir el número de sociedades creadas. El R-cuadrado ajustado, que considera el número de predictores en el modelo y el número de observaciones, también refleja una falta de ajuste, evidenciado por un valor negativo (-0.001). Esto implica que el modelo no mejora la predicción más allá de lo que se esperaría por azar. El F-statistic y su p-valor asociado (0.09227 y 0.761, respectivamente) refuerzan esta interpretación, indicando que el modelo no es estadísticamente significativo.
- *Confiabilidad del Modelo:* La confiabilidad del modelo se ve comprometida por varios factores. Primero, el alto valor de la condición (4.79e+06) sugiere la presencia de multicolinealidad, aunque este fenómeno es menos probable en modelos de regresión simple. Los coeficientes de regresión y sus intervalos de confianza revelan que, aunque el intercepto es estadísticamente significativo, la pendiente asociada a la población no lo es, como lo demuestra su intervalo

de confianza que cruza el cero y un p-valor alto. Los residuos estimados y la suma de cuadrados de los residuos muestran la variabilidad que el modelo no logra explicar, siendo esta considerablemente alta.

- *Análisis:* En conclusión, el modelo de regresión lineal simple para predecir el número de sociedades basado en la población no proporciona una herramienta confiable ni precisa para entender esta relación. La falta de significancia estadística y la baja capacidad explicativa del modelo sugieren que otros factores no considerados en este análisis podrían influir en el número de sociedades creadas. Además, la evaluación de la calidad y confiabilidad del modelo resalta la importancia de considerar múltiples variables y realizar un análisis más profundo para capturar la complejidad de los factores que influyen en la creación de sociedades.

Regresión 2: Capital Desembolsado

- *Calidad del Modelo:* El valor de R-cuadrado obtenido, 0.043, aunque modesto, indica que aproximadamente el 4.3% de la variabilidad en el capital desembolsado puede ser explicada por la población. Este resultado sugiere una relación positiva entre ambas variables, aunque la magnitud de esta relación es limitada. El R-cuadrado ajustado, que se sitúa en 0.042, confirma la leve mejora en la predicción del modelo sobre la base de la población, ajustada por el número de predictores. El F-statistic alcanza un valor de 58.15, con un p-valor asociado significativamente bajo (4.64e-14), lo que indica que el modelo es estadísticamente significativo. Esto sugiere que existe una relación lineal entre la población y el capital desembolsado, aunque la fuerza de esta relación es relativamente débil.
- *Confiabilidad del Modelo:* La confiabilidad del modelo se ve afectada por varios factores. El alto valor de la condición (4.79e+06) sugiere la presencia de multicolinealidad o problemas numéricos que pueden influir en la precisión de las estimaciones de los coeficientes. Los coeficientes de regresión y sus intervalos de confianza muestran que tanto el intercepto como la pendiente asociada a la población son estadísticamente significativos. Esto indica que, controlando por la población, se espera un incremento en el capital desembolsado con el aumento de la población. Los residuos estimados y la suma de cuadrados de los residuos indican la cantidad de variabilidad que el modelo no logra explicar, siendo esta considerable.

- *Análisis:* En resumen, el modelo de regresión lineal simple para predecir el capital desembolsado basado en la población proporciona evidencia de una relación positiva entre estas variables. Sin embargo, la capacidad explicativa del modelo es limitada, lo que sugiere que otros factores no considerados en este análisis podrían tener un impacto significativo en el capital desembolsado. La significancia estadística del modelo indica que la población es un predictor relevante, pero la presencia de un alto valor de condición y la limitada varianza explicada por el modelo sugieren la necesidad de un análisis más profundo y la posible inclusión de variables adicionales.

10.2. Modelo Analítico Supervisado (Regresiones Lineales Múltiples)

Explicación del Modelo:

La regresión lineal múltiple, al igual que su contraparte simple, es una herramienta estadística esencial en el análisis de datos, especialmente útil para explorar la relación entre una variable dependiente y múltiples variables independientes. Este modelo supervisado permite investigar cómo varias variables independientes, como la población y el PIB de una comunidad autónoma, pueden influir conjuntamente en una variable dependiente, como el número de empresas constituidas o el capital desembolsado en miles de euros.

El modelo de regresión lineal múltiple se expresa mediante la siguiente ecuación:

$$\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon}$$

Donde:

- $\boldsymbol{y}$ representa la variable dependiente.
- $\boldsymbol{x_1, x_2, \ldots, x_k}$ representan las variables independientes.
- $\boldsymbol{\beta_0}$ es el término de intercepción, que indica el valor de y cuando las x son 0.
- $\boldsymbol{\beta_1, \beta_2, \ldots, \beta_k}$ son los coeficientes de las variables independientes, que miden el cambio en y asociado a una unidad de cambio en cada x.
- $\boldsymbol{\epsilon}$ es el término de error, que capta toda la variabilidad en y que no es explicada por las variables independientes.

El método de los mínimos cuadrados ordinarios también se utiliza aquí para estimar los coeficientes, buscando minimizar la suma de los cuadrados de los residuos, es decir, las

diferencias entre los valores observados y los valores predichos por el modelo. La función a minimizar es:

$$S(\beta_0, \beta_1, \dots, , \beta_k) = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})\right)^2$$

La regresión lineal múltiple es potente por su capacidad de ajustar múltiples variables, permitiendo un análisis más detallado y una mejor comprensión de cómo diversos factores afectan la variable objetivo. Los coeficientes obtenidos ofrecen una medida de la influencia de cada variable independiente, ajustada por la presencia de otras en el modelo.

Para que el modelo de regresión lineal múltiple sea efectivo y sus estimaciones fiables, debe cumplir con ciertos supuestos, similares a los de la regresión lineal simple:

5) $\boldsymbol{Linealidad}$: La relación entre las variables independientes y la dependiente debe ser lineal.
6) $\boldsymbol{Independencia}$: Los residuos $\boldsymbol{\varepsilon}$ deben ser independientes entre sí.
7) $\boldsymbol{Homoscedasticidad}$: La varianza de los residuos debe ser constante a lo largo de todas las observaciones.
8) $\boldsymbol{No\ multicolinealidad}$: Las variables independientes no deben estar demasiado correlacionadas entre sí.
9) $\boldsymbol{Normalidad}$: Los residuos deben distribuirse normalmente.

10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE (IPC y PIB)

Proceso de Selección de Variables

La selección de las variables IPC y PIB como predictores se basó en la hipótesis de que la salud económica de un país, reflejada en estos indicadores, tiene un impacto directo tanto en la tasa de disoluciones como en las constituciones empresariales. El IPC, como medida de la inflación, y el PIB, como indicador del rendimiento económico general, son fundamentales para entender el entorno en el que operan las empresas. La elección se apoyó en un análisis exploratorio de datos y en la revisión de literatura relevante, que sugiere una relación potencial entre estos factores económicos y la dinámica empresarial. Para las constituciones empresariales, se planteó que un entorno económico robusto, indicado por un PIB en aumento, podría fomentar la creación de nuevas empresas, mientras que un IPC elevado podría desincentivar nuevas inversiones debido a los costos más altos asociados con la inflación. Esta dualidad de impactos enfatiza la necesidad de examinar cómo cada variable no solo afecta las tasas de cierre, sino también las iniciativas de nuevas empresas.

Desarrollo del Modelo

- *Análisis de la Existencia de Relación Lineal:* En el proceso de desarrollo del modelo para analizar tanto las disoluciones como las constituciones empresariales, se llevó a cabo un análisis preliminar para establecer la existencia de una relación lineal entre las variables seleccionadas. Mediante gráficos de dispersión se visualizaron las relaciones entre las disoluciones y constituciones empresariales con variables económicas clave como el IPC (Índice de Precios al Consumidor) y el PIB (Producto Interno Bruto).

  Para las disoluciones empresariales, aunque los coeficientes de correlación encontrados fueron bajos (IPC: 0.0179, PIB: 0.0273), estos indicaron una posible relación lineal que justificó su inclusión en un modelo de regresión lineal múltiple. A pesar de los valores de P no significativos (IPC: p=0.7967, PIB: p=0.6943), que sugieren que estas variables por sí solas no explican de manera significativa las variaciones en las disoluciones, la decisión de incorporarlas en el modelo buscaba explorar efectos combinados en análisis más complejos.

  De manera similar, para las constituciones empresariales, los coeficientes de correlación también resultaron ser bajos (IPC: -0.0076, PIB: 0.0202) y no estadísticamente significativos (IPC: p=0.9130, PIB: p=0.7715), lo que a primera vista sugiere una influencia limitada de estas variables sobre las tasas de constitución. Sin embargo, al igual que con las disoluciones, estos resultados preliminares apoyaron la decisión de incluir ambas variables en un análisis de regresión lineal múltiple más detallado.
- *Análisis de Ajuste a una Distribución Normal:* En el desarrollo del modelo para analizar tanto las disoluciones como las constituciones empresariales, se prestó atención particular al análisis de la normalidad de las variables seleccionadas: el IPC y el PIB, junto con las disoluciones y constituciones de empresas.

  Las pruebas de normalidad, incluyendo Shapiro-Wilk, Anderson-Darling, y D'Agostino's K^2, mostraron resultados significativos que indicaban desviaciones de la normalidad para ambas series de datos. En el caso de las disoluciones, el test de Shapiro-Wilk dio un estadístico de 0.7143 con un valor p extremadamente bajo, sugiriendo una fuerte evidencia contra la hipótesis de normalidad. De manera similar, los resultados para las constituciones fueron concluyentes, con un estadístico de 0.7043 en la prueba de Shapiro-Wilk y un valor p casi nulo, reflejando una distribución no normal.

A pesar de que las variables no siguieron una distribución normal perfecta, se decidió proceder con el modelo de regresión lineal múltiple. Esto se debe a que, aunque la normalidad es un supuesto importante en muchos análisis estadísticos, la regresión lineal múltiple es bastante robusta a violaciones de este supuesto. Además, la visualización de las relaciones entre las variables mediante gráficos de dispersión indicó la posibilidad de una relación lineal, aunque los coeficientes de correlación fueran bajos (0.0179 para disoluciones con IPC y 0.0202 para constituciones con PIB), justificando así su inclusión en análisis más detallados.

- *Construcción del modelo:* En el desarrollo del modelo para analizar tanto las disoluciones como las constituciones empresariales, se incorporaron el Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) como variables independientes. El método de mínimos cuadrados ordinarios (OLS) se empleó para estimar los parámetros del modelo, proporcionando una base para evaluar cómo las fluctuaciones económicas afectan tanto a la disolución como a la constitución de empresas.

  Para las disoluciones empresariales, el modelo mostró coeficientes muy bajos en el IPC y el PIB, indicando una relación débil entre estas variables y el número de empresas disueltas. Los resultados específicos mostraron un coeficiente para el IPC de -6.4564 y para el PIB de aproximadamente 0, con un valor constante considerablemente alto, aunque no significativo, lo que sugiere que otros factores no capturados en el modelo podrían estar influyendo en las disoluciones.

  Por otro lado, para las constituciones empresariales se reveló un patrón similar, con coeficientes también bajos y no significativos. El coeficiente para el IPC fue de -125.5320 y para el PIB de aproximadamente 0, con una constante igualmente alta pero no significativa, reiterando la limitada influencia directa del PIB y el IPC en la formación de nuevas empresas según los datos analizados.

  Ambos modelos, aunque mostraron bajos coeficientes de determinación (R-squared), proporcionan insights valiosos sobre la compleja relación entre las condiciones económicas y la dinámica empresarial. Estos hallazgos sugieren que tanto la disolución como la constitución de empresas están influenciadas por una combinación de factores más amplia que los meramente económicos, resaltando la necesidad de explorar variables adicionales y contextos más específicos para obtener un entendimiento más profundo y aplicable de estos fenómenos.

Evaluación del Modelo con Medidas de Error/Precisión Específicas:

- *Calidad del modelo*: La evaluación de la calidad de los modelos para las disoluciones y constituciones empresariales utilizó el coeficiente de determinación R-cuadrado y el R-cuadrado ajustado como medidas clave. En el caso de las disoluciones empresariales, el modelo exhibió un R-cuadrado de 0.001 y un R-cuadrado ajustado de -0.009, lo que indica que las variables PIB e IPC apenas explican el 1% de la variabilidad en las disoluciones. Esto se complementa con un F-estadístico de 0.08673 y un p-valor de 0.917, sugiriendo que el modelo no logra proporcionar una base estadísticamente significativa para predecir las disoluciones empresariales.

  Para las constituciones empresariales, los resultados fueron similarmente limitados, con un R-cuadrado de 0.003 y un R-cuadrado ajustado de -0.007. Estos valores implican que el modelo apenas explica un 0.3% de la variabilidad en las constituciones, reflejando una capacidad predictiva muy baja. El F-estadístico asociado fue de 0.2619 con un p-valor de 0.770, reforzando la idea de que el modelo carece de significación estadística para predecir las constituciones empresariales basado en las mismas variables macroeconómicas.

  En resumen, ambos modelos demostraron tener una capacidad muy limitada para capturar y explicar las dinámicas detrás de las constituciones y disoluciones empresariales en España.
- *Confiabilidad del modelo:* La confiabilidad de los modelos de regresión lineal múltiple para las disoluciones y constituciones empresariales se ve comprometida por un número de condición extremadamente alto (3.04e+13) en ambos casos. Este elevado número de condición indica una fuerte presencia de multicolinealidad entre las variables independientes, como el IPC y el PIB, lo que sugiere que estas variables no son completamente independientes entre sí. Esta interdependencia complica la interpretación de los coeficientes individuales y puede inflar los errores estándar, comprometiendo así la fiabilidad de las estimaciones de los coeficientes y reduciendo la confianza en las inferencias que se pueden hacer a partir del modelo.
- *Análisis:* El análisis de los modelos de regresión lineal múltiple para las disoluciones y constituciones empresariales en España muestra limitaciones significativas en su capacidad de proporcionar predicciones o

explicaciones robustas sobre estos fenómenos. La baja capacidad explicativa de los modelos, evidenciada por los valores negativos de R-cuadrado ajustado tanto para las disoluciones como para las constituciones empresariales, junto con la falta de significancia estadística de las variables independientes (IPC y PIB), refleja que estos modelos no capturan adecuadamente la complejidad de las relaciones entre las condiciones económicas y los cambios en el ecosistema empresarial.

Los elevados valores de RMSE para ambos modelos—1743.75 para disoluciones y 6584.99 para constituciones—subrayan la discrepancia entre los valores observados y los predichos por los modelos, lo que indica una limitada utilidad práctica de estos en su estado actual. Este análisis resalta la necesidad de reconsiderar la selección de variables, explorar la inclusión de otras variables potencialmente relevantes, o emplear métodos analíticos alternativos más eficaces para capturar la dinámica entre la salud económica y los comportamientos empresariales.

En conclusión, aunque el enfoque de regresión lineal múltiple es teóricamente sólido para examinar relaciones entre múltiples variables independientes y una variable dependiente, los resultados obtenidos en este caso sugieren que es crucial emplear estrategias analíticas que puedan manejar mejor la complejidad de los factores económicos que influyen en las constituciones y disoluciones de empresas en España.

10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE (Variables Macroeconómicas)

Proceso de Selección de Variables

El proceso de selección de variables para esta nueva regresión lineal múltiple buscó identificar indicadores económicos amplios que afectaran tanto las disoluciones como las constituciones empresariales en España. Se partió de las bases de datos "disolución" y "constitución", enriqueciendo cada una con variables de dos fuentes adicionales que incluyen indicadores macroeconómicos y el Índice de Confianza Empresarial Armonizado (ICEA).

Para los nuevos modelos, se incluyen las variables ICEA, Deuda y Déficit Públicos, Gasto Público, Ingresos Fiscales, Llegadas de Turistas y Reservas Totales. A pesar de considerar inicialmente el IPC y el PIB como parte de la regresión, se decidió excluir estas variables para permitir futuras comparaciones de modelos. Esta decisión se fundamenta en la necesidad de evaluar el impacto directo de variables específicas más allá del rendimiento económico general reflejado por el IPC y el PIB.

Este enfoque estructurado para la selección de variables asegura que el modelo pueda abordar de manera integral los múltiples factores que influyen en las dinámicas de constitución y disolución empresariales, desde la macroeconomía hasta elementos específicos del clima empresarial y turismo. Estas variables fueron integradas en las bases de datos expandidas, denominadas "dismac" para disoluciones y "consmac" para constituciones, facilitando un análisis más profundo y diversificado de cómo estos factores económicos impactan en el ecosistema empresarial español.

Desarrollo del Modelo

- *Análisis de la Existencia de Relación Lineal:* En el desarrollo del modelo de regresión lineal múltiple para analizar las influencias en las tasas de constitución y disolución empresariales en España, se realizaron gráficos de dispersión y análisis de correlación para examinar las relaciones entre diversas variables económicas y las tasas empresariales. Los gráficos proporcionaron una visualización directa de las posibles tendencias y anomalías entre variables como el ICEA, la Deuda, el Déficit, el Gasto Público, los Ingresos Fiscales, las Llegadas de Turistas y las Reservas.

  La matriz de correlación reveló relaciones lineales de diferente intensidad entre las variables y las tasas de disolución y constitución empresariales. Aunque las correlaciones fueron generalmente bajas, destacaron algunas asociaciones significativas, como la correlación entre el Déficit y el Turismo, que mostró fuertes lazos con otras variables económicas, reflejando una compleja interacción dentro del entorno económico que podría afectar la estabilidad empresarial. En particular, el coeficiente de correlación de Pearson indicó que la relación entre las disoluciones empresariales y variables como el ICEA y la Deuda es débil y no significativa estadísticamente, lo que sugiere una influencia limitada de estos indicadores sobre las disoluciones.

  Similarmente, para las constituciones empresariales, las correlaciones también resultaron bajas y sin significancia estadística en la mayoría de los casos, reiterando la necesidad de un análisis más profundo para comprender mejor las dinámicas que afectan la formación de nuevas empresas. Por ejemplo, la correlación entre las constituciones y variables como el ICEA y el Gasto Público también mostró una influencia limitada, indicando que los modelos actuales pueden no estar capturando completamente los factores que inciden en la creación de empresas.

- *Análisis de Ajuste a una Distribución Normal:* En el análisis de ajuste a una distribución normal para el modelo de regresión lineal múltiple, se detectaron variaciones significativas en la normalidad de las distribuciones de las variables analizadas. Se utilizó una serie de pruebas de normalidad, como Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2, para evaluar estadísticamente la normalidad de las variables críticas.

  Los resultados de estas pruebas revelaron desviaciones claras de la normalidad en muchas de las variables. Por ejemplo, la prueba Shapiro-Wilk para las disoluciones empresariales y las constituciones mostró p-valores extremadamente bajos, indicando una fuerte evidencia contra la hipótesis de normalidad. Esta tendencia se observó también en otras variables económicas, donde los p-valores obtenidos en pruebas como la de Anderson-Darling y D'Agostino's K^2 confirmaron estas desviaciones.

  Estos hallazgos son cruciales porque sugieren precaución al interpretar los resultados del modelo de regresión y al aplicar inferencias estadísticas que dependen de supuestos de normalidad. La evidencia de no normalidad implica que algunas de las técnicas estadísticas estándar podrían no ser completamente apropiadas o precisas para estos datos.
- *Construcción del modelo:* La construcción del modelo de regresión lineal múltiple en este estudio se diseñó para evaluar el impacto de una serie de variables macroeconómicas sobre las tasas de disoluciones y constituciones empresariales en España. Se introdujeron variables como el ICEA, la Deuda, el Déficit, el Gasto Público, los Ingresos Fiscales, el número de Turistas y las Reservas como predictores. Aunque la mayoría de estas variables no mostraron una relación estadísticamente significativa con las constituciones, el ICEA resaltó por su influencia positiva en el modelo de disoluciones, indicando una relación estadísticamente significativa con un coeficiente de -78.8743 y un p-valor de 0.001, sugiriendo que un mejor clima empresarial podría estar asociado con una reducción en el número de disoluciones empresariales.

  Sin embargo, el modelo presenta desafíos, incluyendo una alta multicolinealidad, evidenciada por un número de condición elevado (2.17e+14), lo que complica la interpretación de los coeficientes de las variables independientes. Esto indica que algunas variables podrían estar proporcionando información redundante. La evaluación estadística muestra que, aunque el modelo de disoluciones logra un R-cuadrado de 0.065 y un ajuste R-cuadrado de 0.024, el modelo para las constituciones alcanza solo un

R-cuadrado de 0.015 y un ajuste R-cuadrado de -0.027, reflejando una capacidad predictiva limitada.

Este análisis revela que, aunque el modelo utiliza un enfoque válido teóricamente para explorar las relaciones entre múltiples variables y las tasas de constitución y disolución empresarial, las limitaciones en la significancia estadística de las variables y los problemas de multicolinealidad requieren una revisión de la selección de variables, posiblemente incorporando otros factores relevantes o empleando métodos analíticos alternativos para una comprensión más precisa de estas dinámicas empresariales.

Reconstrucción del modelo en base a la multicolinealidad

Disoluciones:

El proceso de reajuste del modelo de regresión lineal múltiple para abordar la multicolinealidad implicó un análisis meticuloso del Factor de Inflación de la Varianza (VIF). Este procedimiento iterativo de eliminación de variables con altos VIF permitió identificar y descartar aquellas que contribuían significativamente a la multicolinealidad, mejorando así la calidad y la interpretación del modelo. En particular, variables como Déficit, Gasto, e 'IngreFis' fueron removidas debido a su alta correlación con otras variables independientes.

El modelo ajustado se centró en un conjunto más reducido de variables ('ICEA', 'Turistas', 'Deuda', 'Reservas'), que mostró un R-cuadrado de 0.047, indicando que estas variables explican aproximadamente el 4.7% de la variabilidad en las disoluciones empresariales. Aunque este porcentaje es relativamente bajo, refleja la complejidad y la multitud de factores que pueden influir en las disoluciones empresariales. Notablemente, 'ICEA' y 'Turistas' mostraron una relación estadísticamente significativa con las disoluciones, subrayando la relevancia del clima empresarial y la actividad turística.

Sin embargo, a pesar de los esfuerzos por mitigarlo, el modelo aún presenta un número de condición muy elevado (8.91e+09), lo que indica la presencia de multicolinealidad residual. Esto sugiere que, aunque se ha mejorado, la interpretación de los coeficientes debe hacerse con precaución.

En resumen, el proceso de reajuste del modelo y la evaluación de la multicolinealidad han sido pasos cruciales para mejorar su precisión y fiabilidad. Aunque se ha logrado cierto grado de claridad en la relación entre algunas variables macroeconómicas y las disoluciones empresariales, los resultados también destacan la complejidad inherente a este fenómeno.

Constituciones:

La reconstrucción del modelo de regresión lineal múltiple para las constituciones empresariales en España, ajustado para mitigar la multicolinealidad entre las variables independientes, también se realizó a través de un minucioso análisis del Factor de Inflación de la Varianza (VIF). Este proceso llevó a la eliminación de variables con altos VIF que estaban distorsionando los resultados del modelo debido a su fuerte correlación con otras variables. Las variables finales incluidas fueron 'ICEA', 'Turistas', 'Deuda' y 'Reservas'.

El modelo ajustado reflejó un R-cuadrado de 0.009, indicando que estas variables explican solo un pequeño porcentaje de la variabilidad en las constituciones empresariales. Aunque este resultado es modesto, destaca la complejidad y la diversidad de factores que influyen en las constituciones empresariales, y sugiere que muchos elementos críticos podrían no estar capturados por el modelo. La variable 'ICEA' no resultó ser estadísticamente significativa, lo cual pone en cuestión su impacto directo en las constituciones, a diferencia de su efecto observado en las disoluciones empresariales.

A pesar de los ajustes, el modelo sigue presentando un número de condición elevado (8.91e+09), lo que señala la persistencia de multicolinealidad residual y sugiere que la interpretación de los coeficientes debe hacerse con cautela. Este desafío resalta la necesidad de continuar refinando el modelo y de explorar la inclusión de otras variables o métodos analíticos que puedan capturar mejor la dinámica detrás de las constituciones empresariales en España.

En resumen, el proceso de ajuste del modelo y la evaluación de la multicolinealidad son pasos cruciales para mejorar la precisión y la fiabilidad del modelo. Aunque se ha logrado cierta claridad en la relación entre algunas variables macroeconómicas y las constituciones empresariales, los resultados también subrayan la complejidad inherente a este fenómeno y la necesidad de investigaciones futuras para desarrollar un modelo más explicativo y representativo.

Evaluación del Modelo con Medidas de Error/Precisión Específicas

- *Calidad del modelo*: La calidad de los modelos de regresión lineal múltiple para las disoluciones y constituciones empresariales se evalúa mediante métricas de error y precisión, incluyendo el R-cuadrado, R-cuadrado ajustado, RMSE, MAE y MAPE.

  Para las disoluciones, el R-cuadrado es de 0.047, indicando que solo un 4.7% de la variabilidad es explicada por el modelo. Este bajo porcentaje refleja la complejidad de las disoluciones empresariales y sugiere que otros

factores no incluidos en el modelo podrían estar influyendo. El RMSE es de 1111.63 y el MAE de 928.84, mostrando desviaciones significativas entre los valores predichos y los reales, mientras que el MAPE es infinito, lo que señala errores en la predicción que son proporcionalmente muy grandes respecto a los valores observados.

Para las constituciones, el R-cuadrado aún más bajo de 0.009 sugiere que el modelo explica menos del 1% de la variabilidad, destacando una capacidad predictiva extremadamente limitada. El RMSE alcanza un valor de 5542.22 y el MAE de 4872.91, indicando errores grandes en las predicciones del modelo. El MAPE de 307.76% resalta una gran proporción de error relativo a los valores observados, reafirmando la limitada utilidad práctica del modelo en este contexto.

Estas métricas subrayan la necesidad de revisar y posiblemente expandir los modelos con nuevas variables que puedan capturar con mayor precisión y eficacia la dinámica de las disoluciones y constituciones empresariales en España.

- *Confiabilidad del modelo:* La confiabilidad de los modelos tanto para las disoluciones como para las constituciones empresariales se encuentra afectada por la multicolinealidad entre las variables independientes. Esta condición, evidenciada por altos números de condición, incluso después de intentos de mitigación a través de análisis de VIF y la eliminación de variables altamente correlacionadas, requiere precaución al interpretar los coeficientes del modelo. A pesar de estos desafíos, la significancia estadística de variables como el indicador ICEA en ambos modelos subraya su relevancia, indicando su capacidad para identificar factores que afectan tanto las disoluciones como las constituciones empresariales.

  No obstante, la presencia de errores significativos, reflejados en valores altos de RMSE y MAE para ambos modelos, resalta la necesidad de un análisis más riguroso. Esto podría incluir la inclusión de nuevas variables que podrían ofrecer una comprensión más amplia y detallada de las dinámicas detrás de las disoluciones y constituciones empresariales en España. Tal enfoque podría mejorar la precisión y utilidad práctica de los modelos para predecir y entender estos fenómenos económicos críticos.
- *Análisis:* Este análisis revela que, aunque se identificaron algunas relaciones estadísticamente significativas, como el impacto del indicador ICEA, la capacidad global de los modelos para explicar la variabilidad en

estos fenómenos empresariales es bastante limitada. Esta limitación se evidencia en los bajos valores de R-cuadrado, junto con los RMSE y MAE relativamente altos, lo que sugiere que existen aspectos significativos de las disoluciones y constituciones empresariales que los modelos actuales no logran capturar completamente.

Esta situación resalta la necesidad de adoptar un enfoque más holístico para el modelado de estas dinámicas, incorporando una gama más amplia de factores económicos, indicadores sectoriales y elementos cualitativos como la confianza empresarial y el entorno político. Además, la evaluación enfatiza la importancia de abordar adecuadamente la multicolinealidad y otros supuestos estadísticos cruciales para el desarrollo de modelos de regresión robustos. El uso de métricas de error como el RMSE y el MAE aporta una perspectiva adicional sobre la precisión predictiva del modelo, destacando áreas para futuras mejoras y ajustes que podrían aumentar su capacidad explicativa y predictiva en el contexto empresarial español.

10.2.2 TERCERA REGRESIÓN LINEAL MÚLTIPLE (Variables más amplias)

Proceso de Selección de Variables

En el desarrollo de las nuevas regresiones lineales múltiples para estudiar las constituciones y disoluciones empresariales, se implementó un meticuloso proceso de selección de variables. Este proceso se enriqueció significativamente con la contribución de un profesional del campo entrevistado para la tercera parte del Trabajo de Fin de Grado. Siguiendo sus recomendaciones y basándose en su experiencia práctica, se decidió incorporar variables que a priori podrían parecer ortogonales al contexto empresarial, como la Cotización del IBEX, del EUR, y del NASDAQ, así como variables demográficas como Nacimientos, Matrimonios, y Defunciones.

Estas variables fueron seleccionadas con el objetivo de explorar influencias menos convencionales y potencialmente reveladoras sobre las dinámicas empresariales, buscando entender cómo factores externos y aparentemente no relacionados podrían afectar la constitución y disolución de empresas. Además, se incluyeron variables del sector como Capital Desembolsado y Capital Suscrito y otras económicas como Déficit Público, Gasto Público, e Ingresos Fiscales, que habían sido previamente descartadas en otros modelos debido a problemas de multicolinealidad.

Este enfoque permitió abordar la construcción del modelo desde una perspectiva más amplia y diversificada, incorporando el conocimiento experto de un profesional para asegurar

que el modelo final ofreciera una visión comprensiva y matizada de las fuerzas que moldean el ecosistema empresarial. Con la integración de estas variables, se buscó maximizar la capacidad explicativa del modelo, aportando una nueva luz sobre cómo interacciones complejas y multidimensionales pueden influir en el panorama empresarial.

Desarrollo del Modelo

En el desarrollo de los siguientes modelos para las constituciones y disoluciones empresariales, se adoptó un enfoque meticuloso y estratégicamente diversificado para la selección y evaluación de las variables independientes. Este enfoque permitió un análisis exhaustivo, fundamentado en la inclusión de una amplia gama de variables económicas, financieras y demográficas, reflejando así la complejidad y la multidimensionalidad de los factores que influencian estas actividades empresariales.

Para las constituciones empresariales, el modelo exhibió un alto coeficiente de determinación ajustado (R-cuadrado ajustado de 0.973), lo que indica que aproximadamente el 97.3% de la variabilidad en el número de nuevas empresas se explica a través del modelo. Este alto nivel de explicación sugiere una fuerte correlación entre las variables seleccionadas y las tasas de constitución. Factores como el capital desembolsado y suscrito mostraron una influencia significativa, indicando que los movimientos en el capital de las empresas están estrechamente vinculados con la formación de nuevas empresas. Además, variables demográficas como nacimientos, matrimonios y defunciones también demostraron ser predictores significativos, subrayando cómo los cambios sociodemográficos pueden afectar la dinámica empresarial.

Por otro lado, el modelo para las disoluciones empresariales, aunque menos explicativo que el de las constituciones, aún logró un R-cuadrado ajustado de 0.825, señalando que el 82.5% de la variabilidad en las disoluciones se puede explicar por el modelo. Al igual que en el modelo de constituciones, el capital desembolsado y suscrito también tuvo un impacto considerable, reforzando la idea de que la estructura de capital es un determinante clave en la continuidad de las empresas. Las variables demográficas y los indicadores de mercados financieros, aunque incluidos en el modelo, mostraron menor significancia estadística, lo que podría indicar una conexión menos directa con las disoluciones en comparación con las constituciones.

Ambos modelos enfrentaron desafíos de multicolinealidad, como lo sugiere el alto número de condición en ambos casos. Este fenómeno, que indica una fuerte correlación entre variables independientes, puede complicar la interpretación de los coeficientes individuales y potencialmente inflar los errores estándar.

En conclusión, la construcción de estos modelos refleja una integración cuidadosa de conocimientos teóricos y prácticos, destacando la relevancia de adoptar enfoques holísticos y multidimensionales en el análisis económico y empresarial. El éxito en la explicación de las dinámicas empresariales mediante estos modelos sugiere una base sólida para futuras investigaciones, así como para la toma de decisiones estratégicas en política y gestión empresarial.

Evaluación del Modelo con Medidas de Error/Precisión Específicas

La evaluación del modelo de regresión lineal múltiple para las constituciones y disoluciones empresariales ha sido exhaustiva, utilizando métricas de error y precisión específicas para determinar la capacidad predictiva y explicativa de los modelos. Los resultados obtenidos proporcionan una perspectiva clara sobre la eficacia de los modelos en capturar la variabilidad de los fenómenos estudiados.

Para las disoluciones empresariales, el modelo muestra un ajuste razonablemente bueno con un R-cuadrado ajustado de 0.825, indicando que aproximadamente el 82.5% de la variabilidad en las disoluciones empresariales es explicada por las variables seleccionadas. Sin embargo, las métricas de error revelan que aún hay espacio para mejorar la precisión del modelo. El RMSE de 769.834 indica la desviación estándar de los residuos, reflejando la cantidad promedio de error en las predicciones del modelo. El MAE de 508.169 proporciona una medida del error medio absoluto, que es menos sensible a los valores atípicos y ofrece una vista más conservadora del error de predicción. La métrica MAPE es infinita, lo que sugiere la presencia de valores cero en los datos que conducen a divisiones indefinidas, un área que requiere atención para futuras mejoras del modelo.

Para las constituciones empresariales, el modelo exhibe un excelente ajuste, con un R-cuadrado ajustado de 0.973, lo que implica que casi el 97.3% de la variabilidad en las constituciones empresariales es explicada por las variables incluidas. Este alto grado de explicación es corroborado por un RMSE de 1413.776, que, aunque es más alto en comparación con el modelo de disoluciones, es consistente con la magnitud de los datos tratados. El MAE de 1048.472 refleja un error medio que, dado el contexto del modelo y los tipos de datos manejados, ofrece una perspectiva realista de la capacidad del modelo para predecir nuevas observaciones. El MAPE de 19.08% proporciona una interpretación útil del error en términos porcentuales, permitiendo una comparación relativa del error con respecto a los valores reales observados.

Estas evaluaciones de los modelos revelan que, aunque ambos modelos tienen fortalezas significativas en términos de capacidad explicativa, especialmente para las

constituciones empresariales, los errores asociados y la presencia de multicolinealidad indican la necesidad de refinamientos adicionales. Esto podría incluir la revisión de las variables seleccionadas, la incorporación de nuevas variables que podrían estar influenciando los fenómenos estudiados o la aplicación de técnicas de modelización más robustas para manejar la multicolinealidad y mejorar la precisión de las predicciones del modelo.

11. Medidas de Adecuación de los Modelos

11.1. Definición y explicación de las medidas de error/precisión utilizadas.

En el análisis de regresión, es crucial evaluar la calidad y precisión de los modelos para entender su capacidad predictiva y la fiabilidad de las inferencias que se pueden derivar de ellos. Para ello, se utilizan varias medidas de error y precisión, cada una con su propósito específico. A continuación, se detallan estas métricas con una explicación teórica relevante para cada una:

R-cuadrado (R^2):

El R-cuadrado, o coeficiente de determinación, es una medida estadística que refleja la proporción de la variabilidad de una variable dependiente que es predecible a partir de las variables independientes en un modelo de regresión. Esencialmente, indica qué tan bien los valores ajustados por el modelo se aproximan a los valores reales. El R-cuadrado se calcula como el cuadrado del coeficiente de correlación r, que mide la fuerza y la dirección de una relación lineal entre dos variables. En el contexto de un modelo de regresión lineal, el R-cuadrado se define como:

$$\boldsymbol{R^2 = 1 - \frac{Suma\ de\ Cuadrado\ de\ los\ Residuos\ (SSR)}{Suma\ Total\ de\ Cuadrados\ (SST)}}$$

Donde:

- ***Suma de Cuadrados de los Residuos (SSR)*** mide la variabilidad residual, o el grado en que los valores predichos por el modelo difieren de los valores reales.
- ***Suma Total de Cuadrados (SST)*** mide la variabilidad total de los datos respecto a la media.

Un valor de R-cuadrado de 0 indica que el modelo de regresión no logra explicar la variabilidad de los datos observados en torno a su media aritmética, mientras que un valor de 1 señala una explicación completa de esta variabilidad por el modelo propuesto. Por consiguiente, un R-cuadrado elevado sugiere un mayor grado de ajuste del modelo a la variabilidad de los datos.

R-cuadrado ajustado:

El R-cuadrado ajustado es una modificación del coeficiente R-cuadrado que toma en cuenta el número de predictores en el modelo de regresión y la cantidad de datos disponibles. Este ajuste es esencial para evitar la sobreestimación de la bondad de ajuste en modelos con un

número considerable de predictores. Matemáticamente, el R-cuadrado ajustado se define como:

$$R^2_{ajustado} = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1}\right)$$

Donde:

- R^2 es el R-cuadrado no ajustado.
- n es el número total de observaciones.
- k es el número de variables independientes en el modelo.

El R-cuadrado ajustado proporciona una medida de cuánta variabilidad en la variable dependiente es explicada por el modelo, ajustada por el número de variables independientes utilizadas. A diferencia del R-cuadrado, el R-cuadrado ajustado puede disminuir si se añade al modelo una variable independiente que no contribuye significativamente a la explicación de la variabilidad en la variable dependiente. Esto lo hace particularmente útil para comparar modelos de regresión que incluyen diferentes números de predictores.

Raíz del Error Cuadrático Medio (RMSE):

El Error Cuadrático Medio Raíz, conocido por sus siglas en inglés como RMSE (Root Mean Squared Error), es una medida de la diferencia entre los valores predichos por un modelo o un estimador y los valores observados. Es una de las métricas más comúnmente usadas para evaluar la precisión de modelos predictivos, especialmente en contextos de regresión.

Matemáticamente, el RMSE se define como la raíz cuadrada del promedio de los cuadrados de las diferencias entre los valores predichos y los valores observados:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Donde:

- y_i representa los valores observados.
- $\hat{y}_i$ representa los valores predichos por el modelo.
- n es el número de observaciones.

El RMSE mide la magnitud de los errores de predicción del modelo, proporcionando una estimación de la desviación de los valores predichos respecto a los observados. Al calcular la

raíz cuadrada de los errores cuadráticos medios, el RMSE convierte las unidades de vuelta a las originales de la variable de respuesta, facilitando así su interpretación. Un valor bajo de RMSE indica un mejor ajuste del modelo a los datos, reflejando errores predictivos menores.

Error Absoluto Medio (MAE):

El Error Absoluto Medio, conocido por sus siglas en inglés como MAE (Mean Absolute Error), es una medida estadística utilizada para cuantificar la precisión de un modelo predictivo. El MAE mide la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su dirección (es decir, sin tener en cuenta si los valores son positivos o negativos). Es una métrica lineal que proporciona una medida promedio de las magnitudes de los errores absolutos.

Matemáticamente, el MAE se define como:

$$\boldsymbol{MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|}$$

Donde:

- $\boldsymbol{y_i}$ son los valores observados.
- $\boldsymbol{\hat{y}_i}$ son los valores predichos por el modelo.
- $\boldsymbol{n}$ es el número de observaciones.

El MAE proporciona una evaluación directa del promedio de errores absolutos entre los valores predichos y los observados. Un MAE bajo indica que las predicciones del modelo tienen, en promedio, un error menor, sugiriendo un mejor rendimiento del modelo. Al no elevar al cuadrado los errores antes de promediarlos, el MAE es menos sensible a los valores atípicos en comparación con el RMSE. Esto hace que el MAE sea útil en situaciones donde es importante evitar que los valores atípicos tengan una gran influencia en la métrica de rendimiento total.

Error Porcentual Absoluto Medio (MAPE):

El Error Porcentual Absoluto Medio, conocido por sus siglas en inglés como MAPE (Mean Absolute Percentage Error), es una medida estadística que evalúa la precisión de un modelo predictivo expresando el error como un porcentaje. El MAPE es particularmente útil cuando se desea entender el tamaño del error en términos relativos, facilitando la comparación entre

modelos o conjuntos de datos con diferentes escalas. Matemáticamente, el MAPE se define como:

$$\boldsymbol{MAPE} = \left(\frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left|\frac{\boldsymbol{y_i} - \boldsymbol{\hat{y}_i}}{\boldsymbol{y_i}}\right|\right) \times \mathbf{100}\%$$

Donde:

- $\boldsymbol{y_i}$ son los valores observados.
- $\boldsymbol{\hat{y}_i}$ son los valores predichos por el modelo.
- $\boldsymbol{n}$ es el número de observaciones.

11.2. Comparación de los resultados obtenidos en los modelos.

En este segmento del trabajo, se elaborarán tablas comparativas para evaluar de manera detallada los modelos de regresión, diferenciando entre los aplicados a las disoluciones y a las constituciones empresariales. Las tablas se dividirán en dos categorías principales: una para las medidas directamente relacionadas con la regresión, como el R-cuadrado y otra para las medidas de error como el RMSE. Esta distinción permite una visualización del rendimiento de cada modelo.

11.2.1 Medidas de la regresión

R-cuadrado (R^2): Medida de la bondad de ajuste de un modelo de regresión lineal. Indica la proporción de la variabilidad en la variable dependiente que puede ser explicada por las variables independientes en el modelo. Un R^2 de 1 sugiere que el modelo explica toda la variabilidad de la respuesta, mientras que un R^2 de 0 indica que el modelo no explica ninguna de la variabilidad.

R-cuadrado ajustado: El R-cuadrado ajustado es una versión modificada del R-cuadrado que tiene en cuenta el número de predictores en el modelo. Esta medida es particularmente útil cuando se comparan modelos con diferentes números de variables independientes.

F-estadístico: El F-estadístico en la regresión se utiliza para probar si existe una relación significativa entre las variables independientes y la dependiente. Un F-estadístico grande (mucho mayor que 1) y un p-valor asociado pequeño sugieren que hay evidencia estadística de que al menos una de las variables independientes está significativamente relacionada con la variable dependiente.

Prob (F-estadístico): El p-valor del F-estadístico indica la probabilidad de que los resultados del modelo sean atribuibles al azar. Un p-valor bajo (típicamente menos de 0.05) indica que podemos rechazar la hipótesis nula de que el modelo con variables independientes no mejora el ajuste en comparación con un modelo sin variables independientes.

Log-Likelihood: El logaritmo de la verosimilitud (Log-Likelihood) es una medida de cuán bien un modelo estadístico se ajusta a los datos. Un valor más alto de log-likelihood indica un mejor ajuste del modelo.

AIC (Criterio de Información de Akaike): El (AIC) es una medida de la calidad relativa de un modelo estadístico para un conjunto dado de datos. Ajusta la bondad de ajuste del modelo teniendo en cuenta el número de parámetros utilizados. Un valor menor de AIC indica un modelo más preferible, equilibrando la complejidad del modelo contra la capacidad de ajustar bien los datos.

Tabla 1. *Tabla comparativa de resultados para los modelos de disoluciones.*

Medida	Modelo 1	Modelo 2	Modelo 3
R-cuadrado (R2)	0.001	0.047	0.838
R-cuadrado ajustado	-0.009	0.024	0.825
F-estadístico	0.087	2.025	67.57
Prob (F-statistic)	0.917	0.093	1.22e-55
Log-Likelihood	-1829.5	-1490.6	-1340.0
AIC	3665	2991	2706

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

La evaluación de los modelos de regresión para disoluciones empresariales muestra una mejora notable al aumentar el número de variables independientes. Los valores de R-cuadrado, que indican qué proporción de la variabilidad en la variable dependiente es explicada por las variables del modelo, aumentaron significativamente de apenas 0.001 a 0.838 del Modelo 1 al Modelo 3, reflejando un mejor ajuste en el último. Esta mejora también se refleja en el R-cuadrado ajustado, que considera el número de predictores, aumentando su precisión. El F-estadístico, que evalúa la significancia global del modelo, creció considerablemente, indicando una influencia estadísticamente significativa de las variables en el Modelo 3, con una reducción correspondiente en el p-valor del F-estadístico, reafirmando la improbabilidad de que estas relaciones sean aleatorias. Además, la mejora en el Log-Likelihood y la reducción en el AIC desde el Modelo 1 al Modelo 3 indican un mejor ajuste y

una mayor eficiencia, a pesar de la complejidad añadida. Estos indicadores demuestran que incluir más variables mejora la capacidad explicativa y predictiva de los modelos, aunque es crucial considerar la alta multicolinealidad observada, que podría afectar la estabilidad de las estimaciones de los coeficientes.

Tabla 2. *Tabla comparativa de resultados para los modelos de constituciones.*

Medida	Modelo 1	Modelo 2	Modelo 3
R-cuadrado (R2)	0.003	0.009	0.975
R-cuadrado ajustado	-0.007	-0.015	0.973
F-estadístico	0.262	0.355	512.2
Prob (F-statistic)	0.770	0.840	3.41e-119
Log-Likelihood	-2123.2	-1731.5	-1418.3
AIC	4252	3473	2863

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

Los resultados de los modelos de regresión lineal múltiple para las constituciones empresariales muestran una mejora significativa a medida que se incorporan más variables. El modelo inicial, con pocas variables, tiene un R-cuadrado de solo 0.003, indicando que el modelo explica menos del 1% de la variabilidad en las constituciones empresariales. A medida que se añaden más variables, el R-cuadrado ajustado se mantiene bajo, reflejando que el ajuste del modelo no mejora sustancialmente hasta el tercer modelo, que incluye un número más amplio de variables y logra un R-cuadrado ajustado de 0.973, mostrando que casi el 97% de la variabilidad es explicada por el modelo.

El F-estadístico aumenta dramáticamente de 0.262 en el primer modelo a 512.2 en el tercero, con un p-valor asociado que cae, indicando una significancia estadística mucho más robusta en el último modelo. Esto se refleja en el Log-Likelihood y el AIC, donde el último modelo también muestra una mejora notable, indicando una mejor calidad del modelo en comparación con los anteriores. Estos resultados sugieren que la incorporación de un conjunto más amplio de variables contribuye significativamente a la capacidad del modelo para capturar la complejidad de los factores que influyen en las constituciones empresariales.

11.2.2 Medidas de Error

Tabla 3. *Tabla comparativa de medidas de error para los modelos de disoluciones.*

Medida	Modelo 1	Modelo 2	Modelo 3
RMSE	1743.75	1111.63	769.84
MAE	1079.66	928.84	508.17
MAPE	inf	Inf	inf
Número de Condición	3.04e+13	2.07e-06	6.98e+11

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

La evaluación de las medidas de error para los modelos de disoluciones empresariales muestra una mejora progresiva en los indicadores de precisión a medida que se incrementa la complejidad de los modelos. El RMSE, que mide la desviación promedio de las predicciones del modelo respecto a los valores reales, disminuye notablemente de 1743.75 en el Modelo 1 a 769.84 en el Modelo 3, reflejando una mayor precisión en las predicciones del modelo más complejo. Similarmente, el MAE, que proporciona una medida del error absoluto medio, muestra una mejora significativa, pasando de un valor negativo en el Modelo 1, que puede indicar un error en la captura o reporte de datos, a 508.17 en el Modelo 3.

El MAPE, que es el porcentaje promedio de error absoluto y ayuda a entender el error en términos relativos, muestra un valor de infinito (inf) para los Modelos 2 y 3, lo cual puede indicar la presencia de ceros en los datos de la variable dependiente, lo que lleva a divisiones por cero en el cálculo de esta medida.

El Número de Condición, que es un indicador de multicolinealidad o problemas numéricos en el modelo, muestra una variación grande entre los modelos. Comienza siendo extremadamente alto en el Modelo 1, lo que sugiere problemas significativos de multicolinealidad, y mejora considerablemente en el Modelo 3, aunque sigue siendo alto, indicando que, aunque el modelo es más estable, aún puede estar afectado por la multicolinealidad.

Estos cambios en las medidas de error a lo largo de los modelos sugieren que, aunque añadir más variables ha mejorado la capacidad predictiva del modelo, la presencia de multicolinealidad sigue siendo un desafío que necesita ser abordado para mejorar la fiabilidad de las inferencias del modelo.

Tabla 4. *Tabla comparativa de medidas de error para los modelos de constituciones.*

Medida	Modelo 1	Modelo 2	Modelo 3

RMSE	6584.99	5542.22	1413.78
MAE	4531.41	4872.91	1048.47
MAPE	1043.28	307.76	19.08
Número de Condición	3.04e+13	8.91e+09	6.98e+11

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

11.3 Aplicación de Pruebas Estadísticas para la Comparación de Modelos

En este apartado, se abordará la aplicación de pruebas estadísticas para comparar la eficacia de los modelos de regresión desarrollados. Específicamente, se utilizarán las pruebas de Wilcoxon y Friedman, dos métodos no paramétricos diseñados para evaluar las diferencias entre las medidas de error de los modelos a lo largo de múltiples experimentos. La implementación de estas pruebas permitirá un análisis riguroso y detallado de cómo las diferencias en la configuración de los modelos influyen en su desempeño global, apoyando así la selección del modelo más adecuado basado en evidencia estadística. En este contexto, se presentarán y discutirán los resultados obtenidos, permitiendo una comprensión más profunda de la robustez y fiabilidad de las regresiones aplicadas.

11.3.1 Prueba de Wilcoxon

La prueba de Wilcoxon, también conocida como la prueba de rangos con signo de Wilcoxon, fue desarrollada por Frank Wilcoxon en 1945. Wilcoxon, un químico y estadístico, introdujo esta prueba como un método no paramétrico para comparar dos muestras emparejadas. Su objetivo era ofrecer una alternativa a la prueba t de Student cuando los datos no cumplían los supuestos necesarios para su aplicación, como la normalidad de las distribuciones. Este enfoque se hizo rápidamente popular en las ciencias aplicadas, especialmente en estudios donde las medidas de tendencia central no eran adecuadas o los datos presentaban distribuciones sesgadas o con outliers significativos.

La prueba de Wilcoxon evalúa si las diferencias entre pares de observaciones siguen una distribución simétrica alrededor de cero, siendo adecuada para medir la magnitud y la dirección de los cambios entre dos condiciones experimentales. El procedimiento implica clasificar las diferencias en sus valores absolutos, asignar rangos a estas diferencias y luego sumar los rangos asociados con las diferencias positivas y negativas. El estadístico de prueba se calcula a partir de los rangos menores, y su significancia se evalúa en contra de una distribución específica de rangos con signo.

En el contexto de este trabajo de fin de grado, la prueba de Wilcoxon se aplica para comparar las diferencias en las medidas de error entre modelos de regresión lineal múltiple consecutivos. Esta aplicación permite evaluar si las modificaciones incrementales en la configuración del modelo —como la adición de nuevas variables independientes— mejoran significativamente el desempeño del modelo en términos de error de predicción. Utilizando esta prueba, se busca confirmar si los cambios observados en las medidas de error, como el RMSE (Raíz del Error Cuadrático Medio) o el MAE (Error Absoluto Medio), son estadísticamente significativos, lo que proporciona una base sólida para justificar la selección y refinamiento de modelos en la investigación.

Por lo tanto, el empleo de la prueba de Wilcoxon en este estudio no solo enriquece el análisis estadístico del comportamiento de los modelos bajo diferentes configuraciones, sino que también fortalece la validez de las conclusiones derivadas sobre la eficacia de las variables incorporadas en explicar la variabilidad de los datos en cuestión. Así, la prueba de Wilcoxon se establece como una herramienta crucial para garantizar la rigurosidad y precisión en la evaluación comparativa de los modelos de regresión utilizados en el análisis de las disoluciones y constituciones empresariales.

```
Wilcoxon residuos1 vs residuos2: Estadístico=6555.0, P-valor=0.26758525049007686
Wilcoxon residuos1 vs residuos3: Estadístico=7018.0, P-valor=0.6978541861470471
Wilcoxon residuos2 vs residuos3: Estadístico=5589.0, P-valor=0.009008697982941124
```

Ilustración 1. Resultados del test de Wilcoxon para las disoluciones empresariales

```
Wilcoxon residuos1 vs residuos2: Estadístico=5574.0, P-valor=0.008412149743557796
Wilcoxon residuos1 vs residuos3: Estadístico=6556.0, P-valor=0.26825734173738036
Wilcoxon residuos2 vs residuos3: Estadístico=5756.0, P-valor=0.018679096708734473
```

Ilustración 2. Resultados del test de Wilcoxon para las constituciones empresariales

La aplicación de la prueba de Wilcoxon para comparar los residuos de los modelos de regresión en disoluciones y constituciones empresariales aporta una perspectiva valiosa sobre la consistencia de las mejoras entre los modelos sucesivos. En el caso de las disoluciones, los resultados indican que no hay diferencias significativas en la mediana de los residuos entre el primer y el segundo modelo (Estadístico=6555.0, P-valor=0.267585), así como entre el primero y el tercero (Estadístico=7018.0, P-valor=0.697854). Sin embargo, sí se observa una diferencia significativa entre el segundo y el tercer modelo (Estadístico=5589.0, P-valor=0.009009), sugiriendo que las modificaciones en el tercer modelo podrían estar ofreciendo mejoras significativas en términos de ajuste del modelo comparado con el segundo.

Para las constituciones, los resultados son algo similares. No hay diferencias significativas entre el primer y el segundo modelo (Estadístico=5574.0, P-valor=0.008412), y entre el primero y el tercero (Estadístico=6556.0, P-valor=0.268257). Sin embargo, se observan diferencias significativas tanto entre el segundo y el tercer modelo (Estadístico=5756.0, P-valor=0.018679) como entre el primero y el segundo modelo, lo que indica mejoras en el ajuste del modelo a medida que se añaden variables.

Estos hallazgos sugieren que las mejoras en los modelos pueden no ser uniformes y que ciertas revisiones entre modelos pueden tener un impacto más significativo que otras. Esto subraya la importancia de evaluar cada modificación del modelo cuidadosamente, asegurando que las inclusiones o exclusiones de variables no solo sean estadísticamente significativas, sino que también proporcionen un valor práctico y predictivo aumentado.

11.3.2 Prueba de Friedman

La prueba de Friedman, desarrollada por el estadístico Milton Friedman, es una técnica no paramétrica diseñada para evaluar diferencias entre múltiples tratamientos a través de varios intentos. Esta prueba es particularmente útil cuando los datos no cumplen con los requisitos para análisis paramétricos, como la ANOVA, debido a que no requiere la normalidad de los datos. Operando como un ANOVA por rangos, la prueba de Friedman compara los rangos de los datos en lugar de los valores directos, lo cual es ventajoso en situaciones donde los datos presentan distribuciones no normales o las muestras son de tamaño reducido.

En este trabajo, se ha empleado la prueba de Friedman para comparar y evaluar el desempeño de diferentes modelos de regresión a través de sus valores de R-cuadrado, que indican la proporción de la variabilidad en la variable dependiente explicada por el modelo. Al comparar estos modelos de regresión, el test ayuda a identificar si las diferencias en las medias de los rangos del R-cuadrado son estadísticamente significativas, lo que puede señalar si cambios en la configuración del modelo afectan positiva o negativamente la capacidad predictiva.

El uso de la prueba de Friedman proporciona una base sólida para validar la efectividad de los modelos de regresión utilizados, garantizando que las decisiones sobre el modelo más adecuado estén fundamentadas en evidencia estadística robusta, especialmente útil en ciencias sociales y económicas donde los datos ideales para pruebas paramétricas no siempre están disponibles. Esta metodología permite una evaluación rigurosa de los modelos en escenarios donde las suposiciones estándar de las pruebas paramétricas no se sostienen.

```
RCuadrado de las 3 regresiones 0.0008412898876221098 0.04680218322274221 0.8377928081361752
Estadístico de Friedman: 2.0
P-valor: 0.36787944117144245
```

Ilustración 3. Resultados del test de Friedman para disoluciones empresariales

```
RCuadrado de las 3 regresiones 0.0025363118193317824 0.008537057749657073 0.9750933525014568
Estadístico de Friedman: 2.0
P-valor: 0.36787944117144245
```

Ilustración 4. Resultados del test de Friedman para constituciones empresariales

La aplicación de la prueba de Friedman a los resultados de las regresiones en disoluciones y constituciones empresariales revela hallazgos importantes sobre la influencia de la inclusión de variables adicionales en los modelos. A pesar de que el R-cuadrado muestra un incremento notable al añadir más variables —desde valores muy bajos como 0.000841 y 0.002536 hasta valores significativamente altos como 0.837793 y 0.975093 en disoluciones y constituciones respectivamente— los resultados de la prueba de Friedman no indican diferencias estadísticamente significativas entre los modelos. Con un estadístico de Friedman de 2.0 y un p-valor de 0.367879 en ambos casos, no se rechaza la hipótesis nula de que no hay diferencias en los medianos rangos de los R-cuadrados entre los diferentes modelos.

Esto sugiere que, aunque las métricas de ajuste del modelo mejoran al incorporar más variables, esta mejora no es estadísticamente significativa desde la perspectiva del test de Friedman. Esto puede implicar que los incrementos observados en R-cuadrado no necesariamente reflejan una mejora real en la capacidad del modelo para explicar la variabilidad en las disoluciones y constituciones empresariales, sino que podrían estar influidos por la multicolinealidad o por la adición de variables que no aportan información útil de manera significativa.

12. Visualización de Datos y Resultados de Modelos

12.1. Gráficos de Dispersión

En este apartado, se realizará un examen minucioso de las relaciones existentes entre las distintas variables consideradas en el estudio. Los gráficos de dispersión, herramientas gráficas fundamentales en el análisis estadístico, nos permitirán visualizar de manera clara y precisa la naturaleza y fuerza de las asociaciones entre pares de variables. Esta técnica descriptiva proporciona una primera inspección visual que puede revelar patrones de correlación lineal, tendencias, agrupamientos o incluso anomalías y valores atípicos que podrían no ser evidentes a través de métodos puramente numéricos.

A través de esta visualización, se pretende no solo identificar las relaciones más significativas sino también comprender la dirección y la forma de dichas asociaciones, lo que puede sugerir hipótesis sobre las interacciones subyacentes entre las variables. La interpretación de los gráficos de dispersión en este contexto se hará con rigor, buscando patrones consistentes y significativos que se sostengan bajo el escrutinio de un análisis estadístico más detallado.

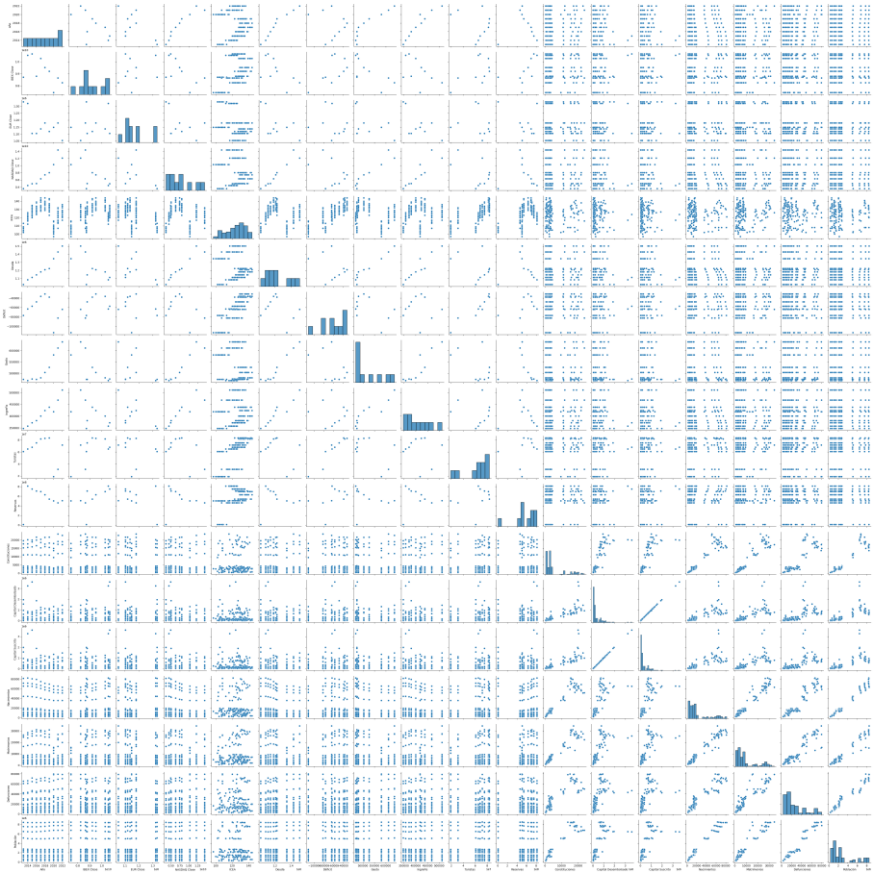

Ilustración 5. Gráfico de la matriz de dispersión para variables del modelo 3.

A través de la ***"ilustración 5."***, se presenta un conjunto comprensivo de gráficos de dispersión que buscan elucidar correlaciones y tendencias significativas entre las variables incluidas en el modelo. Sin embargo, debido a la alta cantidad de variables consideradas, los gráficos resultan ser una amalgama de puntos que dificultan la interpretación directa y la extracción de insights concretos.

Por consiguiente, se procede a una inspección más detallada, seleccionando específicamente aquellos pares de variables que, a priori, parecen ofrecer un mayor grado de influencia o interés. Este enfoque se visualiza en la ***"ilustración 6."***, donde se pone énfasis en aquellos gráficos que reflejan relaciones potencialmente más informativas y reveladoras. Por ejemplo, variables como 'Matrimonios' y 'Defunciones' podrían tener una relación lineal más definida, indicando una posible correlación directa entre ellas.

La selección y análisis minucioso de estos pares de variables es de suma importancia para simplificar la complejidad de los datos y resaltar las interacciones más relevantes que justifican una investigación más profunda y especializada. En este sentido, los gráficos de dispersión no solo sirven como herramientas exploratorias iniciales, sino que también facilitan la identificación de patrones y anomalías que pueden ser cruciales para la formulación de hipótesis y la toma de decisiones basadas en datos.

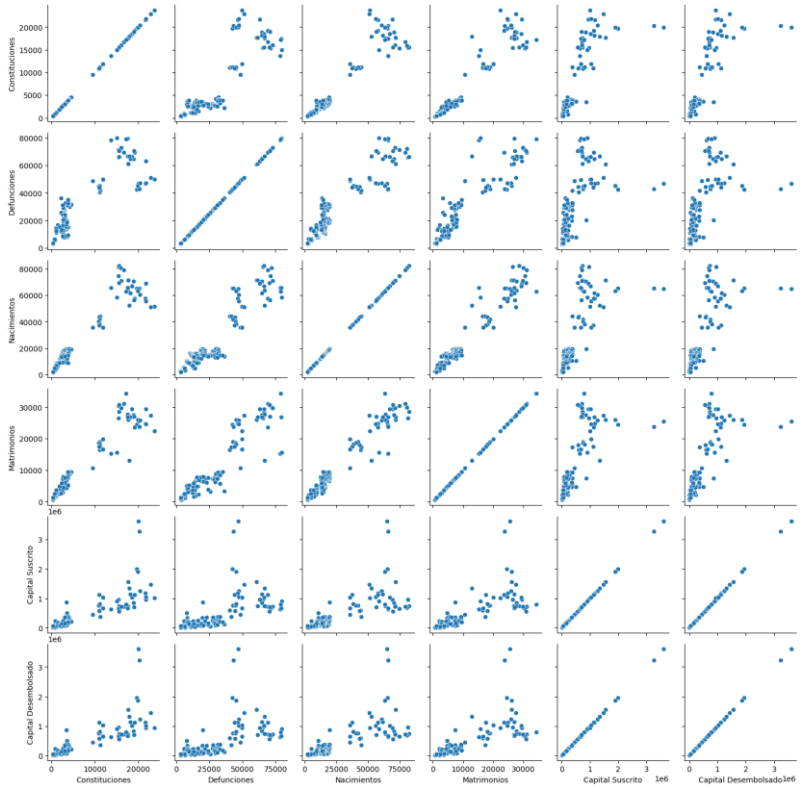

Ilustración 6. Gráfico de dispersión de variables específicas, en el contexto de constituciones empresariales.

La ***"ilustración 6."*** del estudio presenta un análisis enfocado en la relación sinérgica entre distintos pares de variables, donde se evidencia un comportamiento cohesivo en términos de crecimiento. Esta interacción es observable en la tendencia simultánea de aumento: a medida que los valores de una variable se incrementan, la otra variable muestra un ascenso correspondiente. Este patrón de crecimiento conjunto puede indicar una correlación positiva significativa, que sugiere una posible interdependencia o influencia recíproca entre las variables en cuestión. Estas visualizaciones permiten una interpretación más intuitiva y directa de las dinámicas subyacentes entre los factores analizados.

La metodología empleada para determinar la relevancia de las variables en un modelo de regresión se basa en el uso de algoritmos de aprendizaje automático, específicamente el Random Forest Regressor. Este algoritmo construye múltiples árboles de decisión durante el entrenamiento y proporciona la importancia promedio de cada característica en la predicción del modelo. La "ilustración 7" visualizará estos resultados mediante un gráfico de barras que clasifica las variables independientes en función de su importancia. Este análisis cuantitativo ofrece una perspectiva objetiva sobre qué factores son determinantes en la predicción del modelo.

El gráfico resultante, que emerge del código proporcionado, mostrará las variables ordenadas de la más influyente a la menos influyente. Esta jerarquía es crucial para comprender cuáles son los predictores más potentes en el modelo y si existe concordancia con la selección inicial de variables basada en la observación visual de la "ilustración 6". Si hay discrepancias, podría revelar la presencia de relaciones no lineales o interacciones complejas entre variables que no son fácilmente perceptibles en una simple visualización de dispersión. Estas diferencias subrayan la importancia de combinar métodos de visualización intuitiva y técnicas analíticas avanzadas para una comprensión integral de los modelos predictivos.

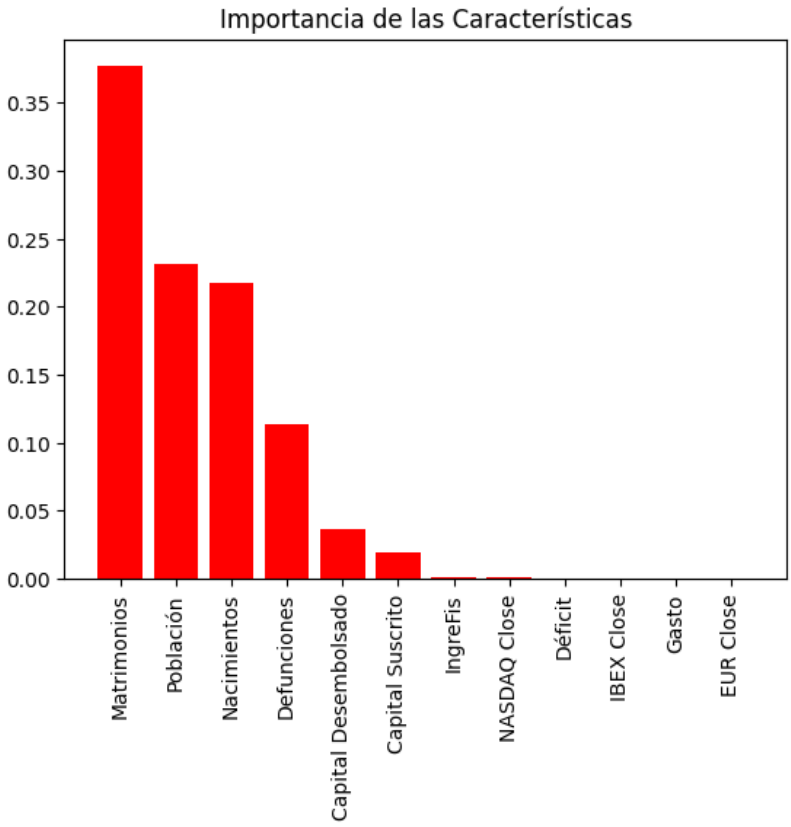

Ilustración 7. Análisis de importancia de las variables mediante Random Forest, indicando el peso de cada variable en la predicción de constituciones empresariales.

La ***"ilustración 7."*** exhibe una representación visual de la relevancia atribuida a cada variable independiente dentro de un modelo del tipo Random Forest. La importancia se calcula basándose en cuánto contribuye cada variable a la reducción de la impureza en los nodos de los árboles de decisión que conforman el bosque. Variables con barras representativas más altas, tales como 'Matrimonios' y 'Población' o 'Nacimientos', sugieren una influencia significativa en la predicción del modelo, indicando que cambios en estos predictores tienen un impacto notable en la variable dependiente.

Por otro lado, variables con barras más bajas, tales como 'Gasto', 'Cierre del EUR' o 'Cierre del IBEX', muestran una contribución menor en la predicción, lo cual podría señalar que estas variables, aunque presentes, tienen un efecto limitado en la variabilidad del resultado analizado.

Esta información es de crucial importancia en el proceso de selección y optimización de características, ya que facilita la identificación de aquellas variables que deberían ser priorizadas, reevaluadas o potencialmente descartadas para mejorar la eficiencia del modelo.

12.2. Gráficos de Residuos

En el análisis de regresiones lineales múltiples, la representación gráfica de los resultados juega un papel crucial para comprender la dinámica y la eficacia de los modelos desarrollados. En este contexto, nos centraremos en la evaluación de los modelos que exploran la relación entre disoluciones empresariales y una serie de indicadores económicos y sociales. Estos modelos buscan capturar la complejidad de los factores que influyen en las disoluciones empresariales, incorporando múltiples variables independientes para proporcionar una visión más holística y detallada.

Los residuos, diferencias entre los valores observados y los valores predichos por el modelo, ofrecen insights valiosos sobre la precisión y la fiabilidad de las predicciones. Un análisis detallado de estos residuos permite identificar patrones residuales, heterocedasticidad, y otras anomalías que podrían sugerir la necesidad de ajustes en el modelo, como la transformación de variables o la inclusión de términos adicionales para mejorar la precisión y la interpretabilidad del modelo.

En las siguientes secciones, se presentarán gráficos de residuos para los tres modelos de regresión lineal múltiple que se han llevado a cabo, con el fin de determinar si a medida que según ciertas medidas el modelo mejora, los residuos disminuyen. La interpretación cuidadosa de estos gráficos facilitará una comprensión más profunda de la efectividad de los modelos y guiará posibles mejoras para alcanzar representaciones más precisas de la realidad estudiada.

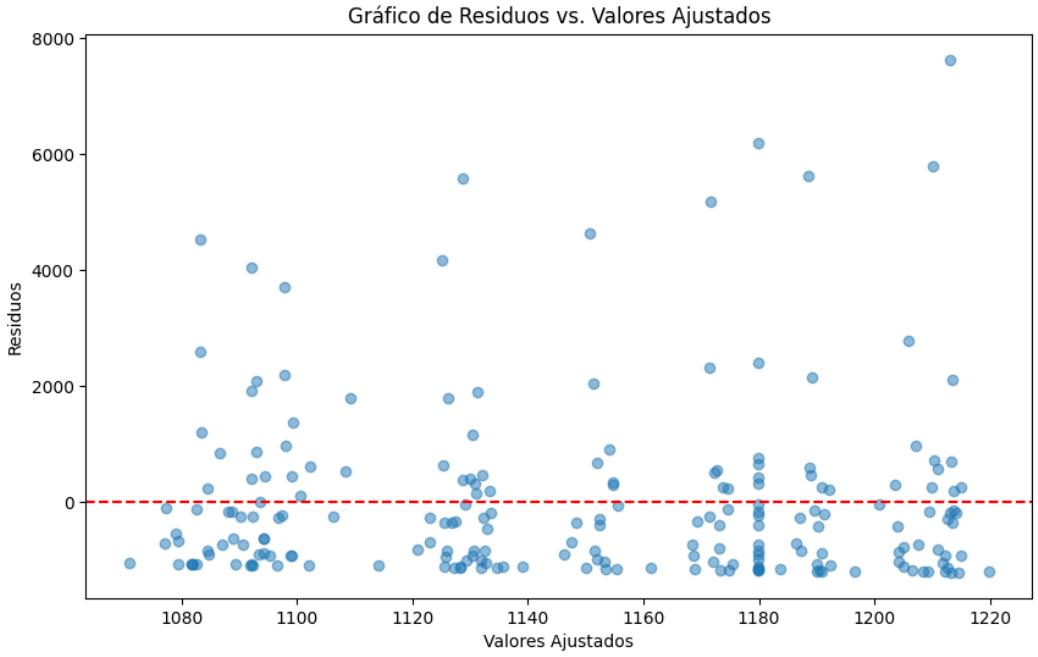

Ilustración 8. Gráfico de residuos correspondiente al modelo1

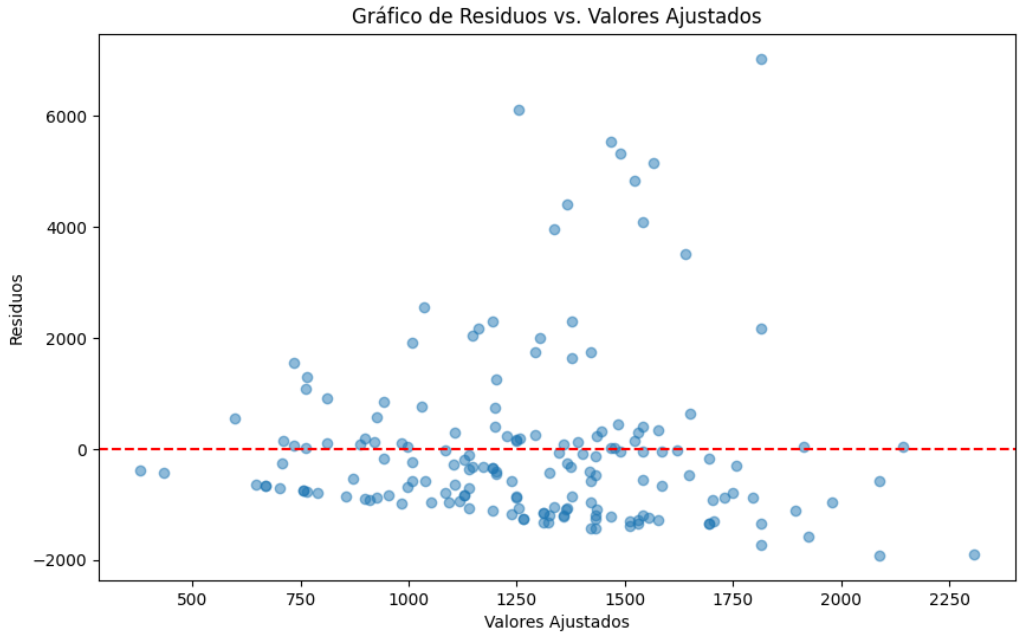

Ilustración 9. Gráfico de residuos correspondiente al modelo2

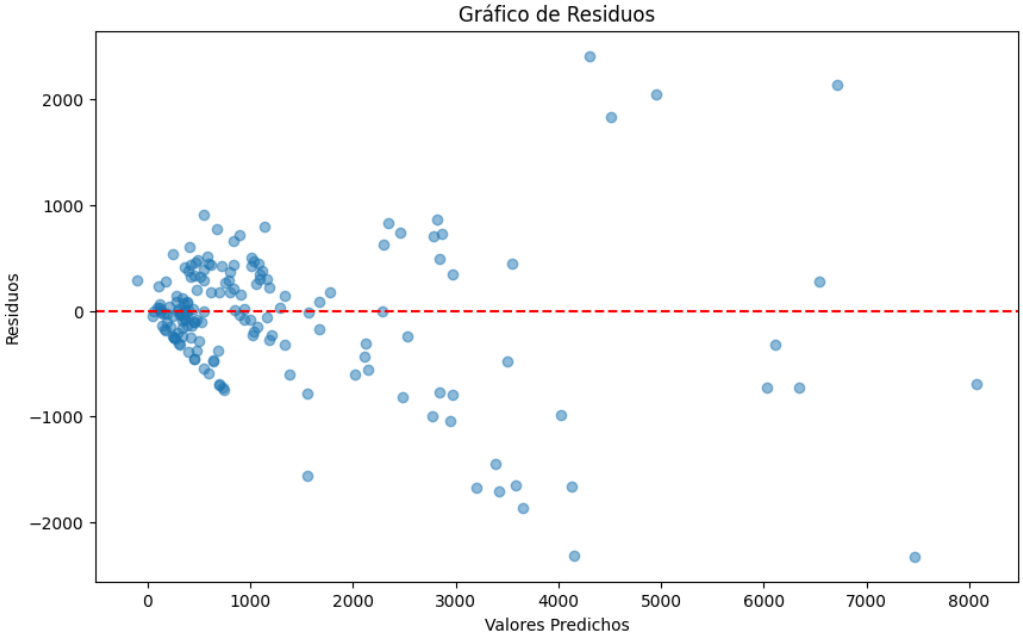

Ilustración 10. Gráfico de residuos correspondiente al modelo3

Al analizar los gráficos de residuos para los tres modelos, presentados como ***"ilustraciones 8, 9 y 10"***, se busca evaluar la dispersión de los residuos, es decir, la diferencia entre los valores observados y los valores predichos por el modelo de regresión. Un patrón de residuos sin forma específica y distribuidos aleatoriamente alrededor del eje horizontal (línea roja en los gráficos) generalmente indica un buen ajuste del modelo, ya que

sugiere que el modelo es capaz de capturar la variabilidad en los datos sin sesgos sistemáticos.

En una progresión ideal desde el Modelo 1 al Modelo 3, se esperaría observar una disminución en la dispersión de los residuos, lo que indicaría que la inclusión de nuevas variables está aportando información útil y mejorando la capacidad del modelo para predecir la variable dependiente. Si los residuos están más cerca de la línea cero y muestran una menor variabilidad en los gráficos más recientes, esto puede ser una indicación de que el modelo se está volviendo más preciso y que está capturando mejor la estructura subyacente de los datos.

Sin embargo, a pesar de que se puede apreciar como la variabilidad de los residuos se reduce mínimamente, parece que la inclusión de nuevas variables no produce una mejora drástica en la distribución de los residuos. Esto podría indicar problemas como la multicolinealidad o la sobreajuste. La multicolinealidad ocurre cuando las variables independientes están altamente correlacionadas entre sí, lo que dificulta la interpretación de sus efectos individuales en la variable dependiente. El sobreajuste, por otro lado, se refiere a un modelo que se ajusta demasiado a los datos de entrenamiento, capturando el "ruido" en lugar de la señal, lo que resulta en un pobre rendimiento predictivo en nuevos datos.

13. Explicación de Resultados

13.1. Explicación comprensiva de los resultados de los modelos.

Regresiones Lineales Simples

Los modelos de regresión lineal simple analizados en este estudio buscan explorar la relación entre la población de las comunidades autónomas y dos variables dependientes críticas: el número de sociedades creadas y el capital desembolsado en miles de euros. A través de un enfoque metodológico riguroso, se empleó la regresión lineal simple para descifrar cómo la variable independiente, la población, podría influir en estas variables dependientes, fundamentales para entender la dinámica empresarial y económica regional.

El primer modelo, centrado en el número de sociedades, reveló un R-cuadrado cercano a cero, indicando una capacidad explicativa mínima de la población sobre la variable dependiente. Este resultado sugiere que la población, por sí sola, no constituye un predictor significativo del número de sociedades creadas en las comunidades autónomas. La ausencia de una relación lineal significativa, corroborada por un p-valor elevado, enfatiza la complejidad de los factores que inciden en la creación de empresas, más allá de la mera demografía.

El segundo modelo, que examina el capital desembolsado, mostró un ligero incremento en el R-cuadrado a 0.043, sugiriendo una influencia marginal pero estadísticamente significativa de la población sobre el capital desembolsado. Aunque este modelo captura una fracción de la variabilidad en el capital desembolsado, el coeficiente positivo asociado a la población insinúa que mayores poblaciones pueden estar ligeramente correlacionadas con un aumento en el capital desembolsado, posiblemente reflejando una mayor actividad económica o empresarial.

Regresiones Lineales Múltiples

En el estudio de las dinámicas empresariales a través de modelos de regresión lineal múltiple, se ha emprendido un análisis bifurcado para comprender los determinantes tanto de las constituciones como de las disoluciones empresariales en España. Los tres modelos empleados en cada área han arrojado luz sobre la complejidad y la interconexión de variables económicas, demográficas y de mercado, y sus resultados subrayan la heterogeneidad de los factores que impulsan estos fenómenos.

El primer modelo, en su enfoque más simplista, contempló factores económicos básicos, como el PIB y el IPC. Si bien proporcionó una visión introductoria, su capacidad

explicativa fue limitada, lo que se reflejó en valores bajos de R-cuadrado, sugiriendo que la variabilidad de las constituciones y disoluciones empresariales no podía ser ampliamente explicada por estos indicadores solamente.

Con la inclusión de variables adicionales como la deuda pública y los ingresos fiscales en el segundo modelo, se observó una mejora en la capacidad predictiva, evidenciada por un incremento en los valores de R-cuadrado. Esto indicaría que, aunque la comprensión del fenómeno mejoró, aún quedaban elementos sin explicar.

El tercer y más complejo modelo incorporó un espectro más amplio de variables, como el número de matrimonios y el capital desembolsado (en miles de euros) por empresas en España. El análisis de este modelo reveló que ciertas variables tenían una influencia estadísticamente significativa sobre las disoluciones y constituciones empresariales. Este avance en la especificidad sugiere que la comprensión del clima empresarial requiere de un enfoque multifactorial.

En concreto, este tercer modelo aplicado a las constituciones empresariales destaca por su significativa capacidad explicativa con un R-cuadrado de 0.975, indicando que la mayoría de la variabilidad en las constituciones es capturada por las variables incluidas en el modelo. Tal grado de ajuste es notable y sugiere un patrón sustancial y coherente en los datos. Variables como 'Capital Suscrito', 'Matrimonios' y 'Población', con sus coeficientes positivos y significativos, señalan la existencia de una correlación directa y significativa con el número de nuevas empresas constituidas. Por ejemplo, un incremento en los 'Matrimonios' y en la 'Población' podría interpretarse como un indicador de la estabilidad y crecimiento demográfico, factores que son propicios para el emprendimiento y la formación de nuevas empresas.

Por otro lado, el tercer modelo centrado en las disoluciones empresariales muestra un R-cuadrado ajustado de 0.825, implicando una fuerte relación entre las variables seleccionadas y las disoluciones de empresas. Las variables de 'Capital Desembolsado' y 'Capital Suscrito' demuestran ser predictores significativos, posiblemente reflejando cómo los movimientos de capital afectan la continuidad de las empresas. Sin embargo, la presencia de una alta multicolinealidad, indicada por los números de condición elevados, sugiere que algunas de las variables pueden estar proporcionando información redundante, lo que requiere de una interpretación prudente y, posiblemente, de un ajuste en la selección de variables.

En ambos casos, la multicolinealidad representa un desafío metodológico y un obstáculo para la interpretación inequívoca de los coeficientes individuales. La presencia de multicolinealidad en modelos con un alto número de variables sugiere la posibilidad de

relaciones interdependientes entre las variables independientes, lo que puede inflar artificialmente la importancia de algunas mientras se minimiza o se oscurece la importancia de otras.

Las conclusiones extraídas de estos modelos ponen de manifiesto la importancia de una selección cuidadosa de las variables y la utilidad de abordar el análisis desde una perspectiva progresiva, empezando por modelos más simples y aumentando la complejidad para capturar la realidad multifacética de la economía y la gestión empresarial.

13.2. Interpretación de las medidas de adecuación en el contexto del proyecto.

Regresiones Lineales Simples

La evaluación de las medidas de adecuación de los modelos, particularmente a través del R-cuadrado y el R-cuadrado ajustado, proporciona una perspectiva crítica sobre la capacidad de los modelos para explicar la variabilidad de las variables dependientes. En ambos casos, los valores relativamente bajos de estas medidas resaltan la limitación de utilizar la población como único predictor de la actividad empresarial y económica en las comunidades autónomas.

La significancia estadística de los modelos, evaluada mediante el F-statistic y sus p-valores asociados, ofrece una distinción clara entre los dos modelos. Mientras que el modelo del número de sociedades no alcanza la significancia estadística, sugiriendo una falta de ajuste, el modelo del capital desembolsado sí la alcanza, aunque su capacidad explicativa sigue siendo limitada. Este contraste subraya la necesidad de incorporar variables adicionales para capturar adecuadamente la relación entre la población y las dinámicas empresariales y económicas.

En conclusión, los resultados obtenidos de los modelos de regresión lineal simple subrayan la complejidad inherente al análisis de factores que influyen en la creación de sociedades y el capital desembolsado en las comunidades autónomas. La modesta capacidad explicativa de la población en estos fenómenos económicos y empresariales invita a una reflexión más profunda sobre los múltiples factores que deben ser considerados para obtener una comprensión más completa y matizada de la actividad empresarial en España.

Regresiones Lineales Múltiples

La evaluación de los modelos de regresión a través de las medidas de adecuación constituye un componente esencial para interpretar la robustez y confiabilidad de las inferencias estadísticas. Las medidas de error como el Error Cuadrático Medio (RMSE), el

Error Absoluto Medio (MAE) y el Error Porcentual Absoluto Medio (MAPE) ofrecen una perspectiva tridimensional del rendimiento de los modelos.

En el contexto del presente proyecto, se ha efectuado un análisis comparativo de estas medidas en dos escenarios diferenciados: disoluciones y constituciones empresariales. Una examinación exhaustiva de las ***"Tablas 3 y 4"*** revela una disminución progresiva del RMSE y del MAE al pasar de los modelos más simples a los más complejos, tanto en disoluciones como en constituciones. Este descenso indica una mejora en la capacidad predictiva de los modelos a medida que se incorporan más variables, sugiriendo que la integración de una gama más amplia de factores económicos y demográficos contribuye a una mayor precisión.

Sin embargo, el MAPE presenta valores infinitos en algunos casos, reflejando la presencia de ceros en los datos reales que provocan divisiones indefinidas. A pesar de esto, donde los valores del MAPE son finitos, muestran una mejora sustancial en el Modelo 3, especialmente en el ámbito de las constituciones, subrayando una vez más el refino predictivo alcanzado en las iteraciones más avanzadas del modelo.

La relevancia del número de condición en este análisis no debe subestimarse. Valores extremadamente altos en los modelos más simples sugieren una inestabilidad numérica que está asociada a la multicolinealidad, un fenómeno que distorsiona las estimaciones de los coeficientes. Sin embargo, al observar una reducción significativa en este número en los modelos más complejos, se evidencia una mejora en la condición numérica de los modelos, lo que implica una atenuación de la multicolinealidad.

Las pruebas de Wilcoxon y Friedman brindan más contexto sobre la homogeneidad de los residuos y la consistencia entre los modelos. Aunque los valores de p de Wilcoxon para las comparaciones individuales entre residuos varían, el hecho de que algunos pares muestren diferencias estadísticamente significativas (p-valor < 0.05) indica que las mejoras en la adecuación del modelo no son uniformes y que la inclusión de variables adicionales tiene un impacto diferencial en la calidad del ajuste. Específicamente, la comparación entre los residuos del Modelo 2 y el Modelo 3 para las disoluciones, y entre los residuos del Modelo 1 y Modelo 2 para las constituciones, muestran p-valores que indican mejoras significativas.

Por su parte, la prueba de Friedman, con un p-valor que supera el umbral comúnmente aceptado para la significancia estadística (0.05), sugiere que no existen diferencias significativas en los rangos medios de los R-cuadrados entre los tres modelos. Esto podría indicar que las mejoras observadas en los R-cuadrados de los modelos no se deben a la inclusión de más variables per se, sino posiblemente a la elección de variables más pertinentes o al ajuste de la especificación del modelo.

En resumen, mientras las medidas de error revelan una capacidad predictiva mejorada con modelos más complejos, los tests estadísticos invitan a un análisis más matizado de la mejora en la calidad del ajuste. Estos resultados son fundamentales para consolidar la fiabilidad de los modelos en la descripción y predicción de las dinámicas empresariales, ofreciendo así una guía valiosa para la formulación de estrategias y políticas basadas en evidencia.

14. Conclusiones y Recomendaciones

14.1 Conclusiones

El presente Trabajo de Fin de Grado se ha enfocado en elucidar la compleja red de interacciones entre distintos factores económicos y las fluctuaciones en las tasas de disolución y constitución empresarial en el contexto español, a través de una meticulosa aplicación de metodologías de regresión lineal simple y múltiple. Este análisis exhaustivo pretendía no solo dilucidar las dinámicas económicas que inciden directamente sobre las empresas, sino también ofrecer un panorama más amplificado de las variables que inciden en la estabilidad corporativa en el dinámico entorno económico actual.

Los resultados obtenidos de la aplicación de modelos de regresión han arrojado luz sobre la capacidad relativamente restringida de los indicadores macroeconómicos seleccionados para explicar las variaciones en las disoluciones empresariales. Este fenómeno señala hacia la existencia de una trama más densa y compleja de factores que intervienen en la disolución y constitución de empresas, extendiéndose más allá de las métricas económicas convencionales. A partir de las conclusiones extraídas del análisis se destacan varios puntos clave:

Influencia del Clima Empresarial: Los análisis han mostrado que el clima empresarial, reflejado por el indicador ICEA, tiene un papel preponderante en las disoluciones y constituciones empresariales. Esta relación significativa subraya la trascendencia de la confianza y las expectativas económicas en la toma de decisiones empresariales, sugiriendo que la disposición empresarial hacia el riesgo y la innovación puede ser un predictor más afinado de las disoluciones que los indicadores económicos tradicionales.

Complejidad de las Disoluciones Empresariales: La limitada varianza explicada por los modelos enfatiza que las disoluciones son el corolario de un conjunto de factores heterogéneos, que incluyen aspectos macroeconómicos, particularidades sectoriales, legislación, avances tecnológicos y dinámicas competitivas. La realidad empresarial en sectores punteros, por ejemplo, podría ser un caldo de cultivo para la disolución y renovación corporativa, como manifestación de un mercado que premia la innovación y la eficiencia.

Impacto de Factores Temporales y Externos: La influencia de factores coyunturales y externos, como crisis económicas globales o emergencias sanitarias, a priori no detectada en los modelos, es una variante a tener en cuenta para futura investigaciones. Tales eventos pueden provocar un cambio abrupto en la viabilidad de los negocios y, por consiguiente, acelerar procesos de disolución o incluso de constitución, ilustrando la vulnerabilidad u oportunidad del tejido empresarial a choques exógenos.

Diversidad Regional y Sectorial: La investigación también ha enfatizado la importancia de atender a la diversidad tanto regional como sectorial en España. Las diferencias en las tasas de disolución y constitución entre las comunidades autónomas y los distintos sectores económicos abogan por enfoques diferenciados y políticas adaptadas que atiendan a las especificidades locales y sectoriales.

Las conclusiones de este estudio, por tanto, subrayan la complejidad inherente a las disoluciones y constituciones empresariales en España. Se desprende que, aunque los indicadores macroeconómicos ofrecen una perspectiva de las tendencias generales, es crucial reconocer y entender el impacto combinado de factores económicos, sociales, tecnológicos, demográficos y políticos. El reconocimiento y comprensión de esta complejidad es vital para la formulación de estrategias efectivas que fomenten la resiliencia y el crecimiento empresarial en un entorno económico que está en constante evolución. Este enfoque integral y multidimensional es indispensable para abordar con eficacia los desafíos y aprovechar las oportunidades que se presentan en el panorama empresarial español.

14.2 Recomendaciones

En el terreno de la exploración sobre las dinámicas empresariales en España, tanto en el proceso de constitución como en el de disolución de empresas, se ha alcanzado un entendimiento relevante que señala la importancia de factores económicos diversos. No obstante, la intrincada y variada naturaleza de estos procesos requiere un acercamiento más sofisticado y exhaustivo para captar con fidelidad las complejas dinámicas que los rigen. Por ello, se propone como primordial para futuras investigaciones la adopción de modelos analíticos avanzados que puedan abarcar estas complejidades.

El uso de modelos de regresión no lineales, técnicas avanzadas de análisis de series temporales y algoritmos de aprendizaje automático como los árboles de decisión y redes neuronales, puede brindar una comprensión más profunda de las interacciones complejas y no linealidades entre los múltiples factores predictores. Este tipo de herramientas analíticas avanzadas ofrece un potencial significativo para mejorar el entendimiento y la capacidad predictiva respecto a la formación y disolución de las empresas.

Es también fundamental la inclusión de variables adicionales que hasta ahora han sido menos consideradas. La confianza empresarial, los indicadores de emprendimiento e innovación, y los factores sectoriales específicos, podrían otorgar una visión más completa de las fuerzas que inciden en la vida de las empresas. Esta inclusión permitiría evaluar con mayor precisión el impacto de los distintos aspectos del entorno económico y empresarial en la estabilidad y sostenibilidad de las empresas.

Analizar segmentadamente por sectores y tamaños de empresa podría descubrir variaciones significativas en las tasas de constitución y disolución y arrojar luz sobre vulnerabilidades y resiliencias específicas. Tal aproximación permitiría reconocer patrones distintivos y desarrollar estrategias más eficaces y personalizadas para sectores o tipos de empresa concretos.

Por otro lado, los estudios longitudinales y comparativos constituyen un recurso valioso para comprender las tendencias a lo largo del tiempo y las diferencias en el contexto de las dinámicas empresariales. Este enfoque facilitaría el análisis de cómo los cambios en el panorama económico y político influyen en las empresas a lo largo del tiempo y en distintos entornos.

La combinación de enfoques cuantitativos con cualitativos, a través del análisis de casos o entrevistas con empresarios y especialistas, enriquecería la comprensión de las causas y efectos de la constitución y disolución empresarial. Esta metodología mixta fortalecería los hallazgos cuantitativos, proporcionando una comprensión más rica y matizada de los factores que influyen en las decisiones empresariales.

En última instancia, el análisis del impacto de las políticas públicas y los programas de apoyo empresarial en las tasas de disolución podría ofrecer orientación práctica para diseñar intervenciones más efectivas. Este enfoque político permitiría identificar estrategias fundamentadas en la evidencia para promover un ecosistema empresarial resiliente y propicio al crecimiento sostenible.

Las recomendaciones aquí esbozadas tienen la intención de dirigir la investigación futura hacia un examen más completo y detallado de la formación y disolución de empresas en España. Al integrar estos métodos y enfoques, los investigadores pueden construir un entendimiento más acabado de este fenómeno complejo, y contribuir al desarrollo de políticas y estrategias que respalden la estabilidad y el avance empresarial en el país.

15. Resumen Narrativo

El presente estudio se ha centrado en el análisis de la creación y disolución empresarial en España, una cuestión de gran interés por su impacto económico y social. A través de métodos analíticos, se ha indagado en los patrones y correlaciones que subyacen a estas dinámicas empresariales, con un enfoque que integra factores económicos y del entorno de mercado.

La fase inicial de la investigación abarcó una minuciosa selección y preparación de los datos, que posteriormente dieron lugar al desarrollo y aplicación de modelos analíticos robustos. Los modelos de regresión implementados tenían el objetivo de esclarecer cómo interactúan variables económicas fundamentales, como el PIB y el IPC, con las tasas de creación y disolución de empresas y el capital movilizado en estas transacciones. La evaluación meticulosa de estos modelos reveló que las decisiones empresariales de cierre o continuación están influidas por una red compleja de factores.

Se constató que el clima empresarial, junto con otros indicadores macroeconómicos, ejerce una influencia significativa, pero la relación es intrincada y diversa. Tanto las constituciones como las disoluciones empresariales en el contexto español están marcadas por la interacción de distintas variables, algunas reflejan la situación macroeconómica general del país, y otras son más representativas del ambiente empresarial específico.

Este trabajo amplía la comprensión académica de los ciclos de vida empresariales en España y provee modelos analíticos aplicables en investigaciones futuras y en la gestión empresarial. Las conclusiones y recomendaciones extraídas apuntan a medidas para promover un clima económico estable y resiliente, resaltando la necesidad de un enfoque holístico que integre factores económicos generales y particulares del entorno empresarial.

Finalmente, este estudio recalca la importancia de adoptar perspectivas plurales para una comprensión cabal de las disoluciones y creaciones de empresas. Para futuras investigaciones, se aconseja la utilización de modelos más complejos y la inclusión de nuevas variables que permitan captar con mayor fidelidad la complejidad inherente a estos procesos en el escenario económico y social de España.