**UNIVERSIDAD NACIONAL AGRARIA LA MOLINA - ESCUELA DE POSGRADO**

**DOCTORADO EN INGENIERÍA Y CIENCIAS AMBIENTALES**

****

**DISEÑO Y ANÁLISIS DE EXPERIMENTOS EN INGENIERÍA Y CIENCIAS AMBIENTALES**

**Actividad:** Práctica

**Docente:**

Ph.D. Christian René Encina Zelada

**Integrantes:**

-Agatha Prado Gárate

-Gustavo De la Cruz Montalvo

-Jhonsy Omar Silva López

-José Augusto Zevallos Ruiz

**Lima – Perú**

**de septiembre del 2024**

**Práctica N° 06:  
Transformaciones estadísticas**

**CONTENIDO:**

[I. INTRODUCCIÓN 2](#_Toc178516646)

[II. OBJETIVO 2](#_Toc178516647)

[III. MARCO TEÓRICO 3](#_Toc178516648)

[IV. METODOLOGÍA 4](#_Toc178516649)

[V. RESULTADOS Y DISCUSIONES 8](#_Toc178516650)

[VI. CONCLUSIONES Y RECOMENDACIONES 12](#_Toc178516651)

[VII. REFERENCIAS BIBLIOGRÁFICAS 13](#_Toc178516652)

[VIII. ANEXOS 14](#_Toc178516653)

# INTRODUCCIÓN

El análisis de datos experimentales en estudios ambientales frecuentemente requiere la aplicación de técnicas que aseguren que se cumplan los supuestos estadísticos para poder interpretar los resultados con validez. Uno de los principales desafíos en la modelización de datos experimentales es la presencia de heterogeneidad de varianzas y la falta de normalidad en los residuos. Estos problemas pueden influir en la precisión y confiabilidad de los análisis estadísticos tradicionales, como el análisis de varianza (ANOVA), que asumen homogeneidad de varianzas y distribución normal de los errores (Osborne, 2010).

Las transformaciones de datos son métodos estadísticos ampliamente utilizados para manejar problemas de heterocedasticidad y para normalizar la distribución de los datos. Entre las transformaciones más comunes se encuentran la transformación logarítmica, la transformación raíz cuadrada y la transformación recíproca. Estas técnicas permiten estabilizar la varianza entre grupos y aproximar los datos a una distribución normal, facilitando así el uso de métodos paramétricos (Box & Cox, 1964). De acuerdo con Osborne (2010), la transformación logarítmica es particularmente efectiva cuando la variabilidad aumenta proporcionalmente con el promedio, mientras que la transformación raíz cuadrada se usa generalmente cuando los datos son recuentos o frecuencias.

En un informe anterior, se exploraron los efectos de las emisiones de óxido nitroso (N₂O) en Sudáfrica bajo diferentes tratamientos de nitrificación. En aquel análisis por bloques, se encontró que los datos de emisión no cumplían con el supuesto de normalidad necesario para la aplicación del ANOVA, lo cual comprometía la validez de los resultados estadísticos. Por ello, se consideró necesario realizar un nuevo enfoque basado en transformaciones de datos para cumplir con los supuestos estadísticos y así mejorar la calidad de los análisis.

En este trabajo, se aplicarán transformaciones a los datos de emisiones de N₂O de los bloques experimentales de fertilización descritos por Smit et al. (2020), con el objetivo de mejorar la normalidad y estabilizar la varianza, lo cual permitirá una evaluación estadística más sólida y la aplicación de modelos paramétricos para la comparación de tratamientos. La transformación de los datos también busca abordar la heterogeneidad observada en las emisiones debido a la variabilidad en la respuesta del suelo y la interacción con los fertilizantes utilizados.

# OBJETIVO

***Objetivo general***

* Aplicar transformaciones estadísticas a los datos de emisión de óxido nitroso (N₂O) provenientes de los bloques experimentales de fertilización en Sudáfrica a partir de los datos del artículo “Grazing under Irrigation Affects N2O-Emissions Substantially in South Africa”, con el fin de cumplir con los supuestos de normalidad y homogeneidad de varianzas, permitiendo un análisis estadístico adecuado mediante ANOVA.

***Objetivos específicos***

* Determinar el tipo de transformación más adecuada (logarítmica, raíz cuadrada u otra) para estabilizar la varianza y mejorar la normalidad de los datos de emisión de N₂O.
* Evaluar el impacto de las transformaciones en los supuestos del ANOVA, incluyendo la normalidad de los residuos y la homogeneidad de las varianzas.
* Comparar los resultados obtenidos antes y después de la aplicación de las transformaciones para determinar su efectividad en mejorar la validez del análisis estadístico.

# MARCO TEÓRICO

**Transformación de datos**

En los análisis estadísticos, la transformación de datos es una herramienta fundamental para mejorar el cumplimiento de los supuestos estadísticos, como la normalidad de los residuos y la homogeneidad de las varianzas (Box & Cox, 1964). La transformación de datos tiene como objetivo principal hacer que los datos sean adecuados para los métodos paramétricos, que a menudo requieren que las observaciones se distribuyan normalmente y que las varianzas sean homogéneas entre los grupos (Osborne, 2010).

Una de las transformaciones más utilizadas es la transformación logarítmica, la cual es particularmente útil cuando la variabilidad de los datos aumenta proporcionalmente con el promedio (Mynbaev, K. T. & Darkenbayeva, G. S., 2019). Esta transformación permite reducir la asimetría de los datos y estabilizar la varianza, siendo adecuada cuando los datos presentan una distribución sesgada hacia la derecha (Field, 2018). La transformación logarítmica se aplica frecuentemente a datos de naturaleza multiplicativa, como aquellos relacionados con tasas de crecimiento o emisiones de gases de efecto invernadero (Smit et al., 2020).

Otra transformación común es la transformación raíz cuadrada, que se utiliza para estabilizar la varianza de datos que representan conteos o frecuencias (Kirk, 2013). Esta transformación es útil cuando los datos contienen valores pequeños o cero, lo cual es característico en estudios de emisiones de gases, donde las mediciones pueden fluctuar considerablemente en función de las condiciones ambientales y los tratamientos (Atkinson, 2002).

La transformación Box-Cox es otra técnica ampliamente recomendada, ya que ofrece una solución flexible que incluye a la transformación logarítmica y a otras posibles transformaciones como casos particulares (Box & Cox, 1964). Esta técnica permite al investigador determinar el parámetro de transformación óptimo que mejor estabilice la varianza y normalice los datos, lo cual es especialmente útil cuando no se tiene conocimiento previo sobre la transformación adecuada.

Sin embargo, aunque las transformaciones pueden ser muy efectivas, también presentan limitaciones. De acuerdo con Carroll y Ruppert (1988), las transformaciones no siempre garantizan que se cumplan todos los supuestos estadísticos, y en algunos casos, es

posible que se necesite recurrir a modelos más robustos o análisis no paramétricos. La transformación de Johnson ha sido sugerida como una alternativa útil cuando la transformación Box-Cox no es suficiente para normalizar los datos, especialmente en estudios con gran variabilidad o presencia de valores atípicos (Johnson, 1949).

**ANOVA (Análisis de varianza)**

Es un método estadístico que se utiliza para comparar las medias de tres o más grupos y determinar si existen diferencias significativas entre ellas. El ANOVA evalúa la variabilidad entre grupos y dentro de los grupos, verificando si las diferencias observadas son superiores a lo esperado por azar (Montgomery, 2013). Es especialmente útil en experimentos con múltiples tratamientos donde se busca identificar si al menos un grupo difiere significativamente del resto.

**Prueba Shapiro-Wilk**

Es un test estadístico utilizado para evaluar la normalidad de los datos. Verifica si una muestra sigue una distribución normal comparando los valores observados con los esperados en una distribución normal. Es una de las pruebas más potentes para detectar desviaciones de la normalidad, especialmente en muestras pequeñas (Razali, N.M., Wah, 2011)

**Prueba de Bartlett**

Es un test estadístico que se utiliza para verificar la homogeneidad de varianzas entre varios grupos. Evalúa si las varianzas de los diferentes grupos son estadísticamente iguales, lo que es un requisito clave para ciertos análisis estadísticos, como el ANOVA. Es particularmente sensible a la normalidad de los datos, por lo que se recomienda su uso cuando se presume que los datos siguen una distribución normal (Snedecor, G.W., Cochran, 1989).

**Prueba de Tukey**

Es un análisis *post-hoc* utilizado tras un ANOVA para realizar comparaciones múltiples entre las medias de los grupos. Su objetivo es identificar cuáles de las diferencias entre los grupos son significativas, controlando el error de Tipo I que podría ocurrir debido a múltiples comparaciones. Es particularmente útil cuando se comparan todos los pares posibles de medias (Hayter, 1984).

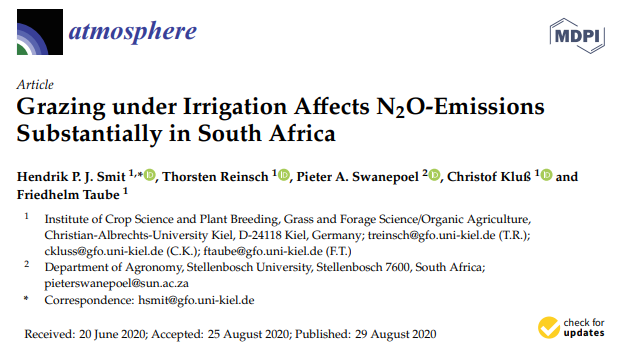
**Prueba de Durbin-Watson**

Es un test estadístico utilizado para detectar la autocorrelación de los residuos en un modelo de regresión lineal. Evalúa si los residuos sucesivos están correlacionados entre sí, lo que violaría el supuesto de independencia en la regresión. Un valor de la prueba cercano a 2 indica que no hay autocorrelación, mientras que valores cercanos a 0 o 4 sugieren una autocorrelación positiva o negativa, respectivamente (Durbin, J., Watson, 1950)

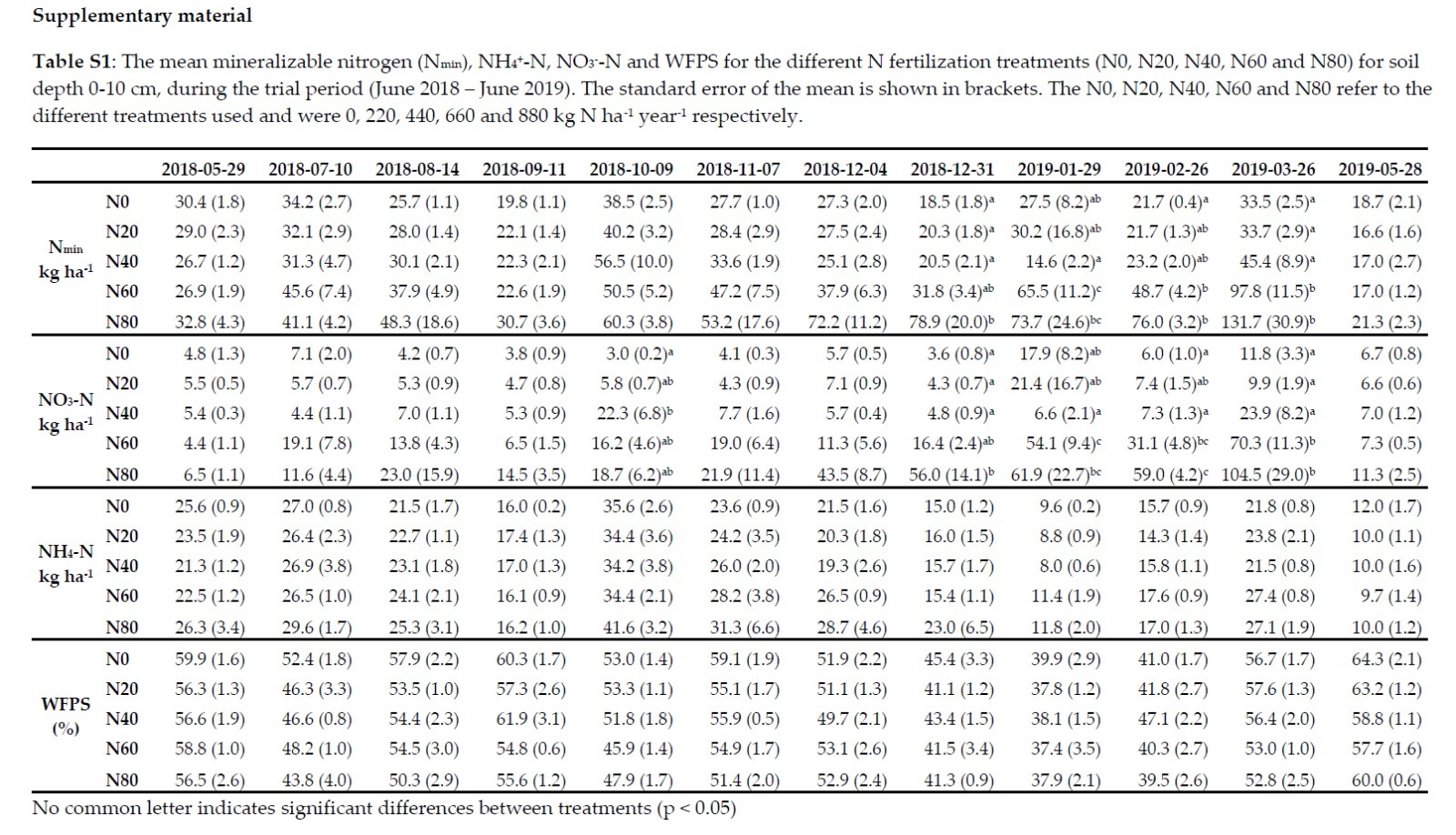
# METODOLOGÍA

***4.1. Descripción del artículo***

El artículo titulado ***"*Grazing under Irrigation Affects N2O-Emissions Substantially in South Africa*"*** examina el impacto de la fertilización y el pastoreo intensivo bajo sistemas de irrigación en las emisiones de óxido nitroso (N2O) en pastos para ganado lechero en Sudáfrica. El estudio evaluó diferentes niveles de fertilización con nitrógeno y utilizó cámaras estáticas para medir las emisiones de N2O durante un año. Los datos suplementarios incluyen información sobre el nitrógeno mineralizable (Nmin), nitrato (NO3-N), amonio (NH4-N), y el contenido de poros llenos de agua (WFPS) para los diferentes tratamientos de fertilización (N0, N20, N40, N60 y N80), obtenidos a una profundidad de 0-10 cm durante el periodo de estudio (junio 2018 - junio 2019). Estos resultados muestran cómo el aumento de la fertilización incrementa las concentraciones de nitrato y amonio en el suelo, lo cual contribuye significativamente a las emisiones de N2O. En el trabajo anterior para el DBCA se utilizó el promedio del nitrógeno mineralizable (Nmin), NH4+-N, NO3--N, pero no el contenido de poros llenos de agua (WFPS) ya que este se encuentra en otra unidad, asimismo se utilizaron los datos de mayo del 2018, tomando como repeticiones los valores extremos hallados a partir de las desviaciones estándar. En el análisis por bloques, se encontró que no se cumplía el supuesto de normalidad para el ANOVA, por lo tanto, en este trabajo aplicaremos transformaciones para cumplir con los supuestos estadísticos

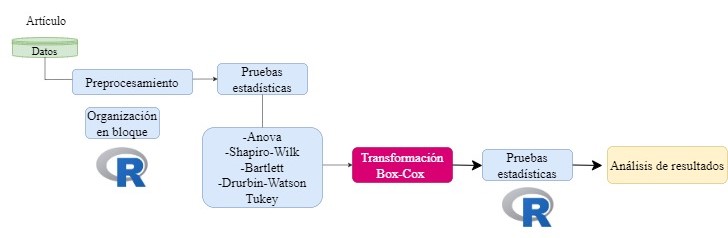


**Figura 1** Datos principales del artículo de Hendrik et al. (2020) ([**https://doi.org/10.3390/atmos11090925**](https://doi.org/10.3390/atmos11090925))



**Figura 2** Datos datos suplementarios del artículo, específicamente sobre el nitrógeno mineralizable (Nmin), nitrato (NO3-N), amonio (NH4-N), y el contenido de poros llenos de agua (WFPS) para diferentes tratamientos de fertilización con nitrógeno (N0, N20, N40, N60 y N80). Estos datos se obtuvieron para el suelo a una profundidad de 0-10 cm durante el periodo de estudio (mayo 2018 - junio 2019).

***4.2. Metodología empleada***



**Figura 3** Flujograma metodológico empleado en el presente trabajo

La Figura 3, el flujo de trabajo de un análisis de datos extraídos de un artículo. Los datos se someten a un **preprocesamiento**, que incluye una **organización en bloque**, seguido de una serie de **pruebas estadísticas** realizadas en el software **R**, tales como ANOVA, Shapiro-Wilk, Bartlett, Durbin-Watson y Tukey. Si los datos no cumplen con los supuestos de estas pruebas, se aplica una **transformación Box-Cox** para ajustar los datos y mejorar su comportamiento estadístico. Posteriormente, se repiten las pruebas estadísticas tras la transformación. Finalmente, el proceso concluye con el **análisis de resultados**, donde se interpretan los hallazgos. Este flujo garantiza que los datos se ajusten adecuadamente a los modelos estadísticos antes de realizar conclusiones robustas.

**4.3. Datos**

La tabla presenta los valores de tres variables de nitrógeno (Nmin, NO3-N y NH4+-N) para cinco tratamientos de fertilización con nitrógeno (N0, N20, N40, N60 y N80). Los tratamientos representan diferentes niveles de aplicación de fertilizante, desde N0 (sin fertilización) hasta N80 (mayor nivel de fertilización). Para el tratamiento N0, los valores de Nmin varían entre 28.6 y 32.2, los de NO3-N entre 3.5 y 6.1, y los de NH4+-N entre 24.7 y 26.5. En el tratamiento N20, los valores de Nmin oscilan entre 26.7 y 31.3, mientras que NO3-N varía de 5 a 6, y NH4+-N de 21.6 a 25.4. El tratamiento N40 presenta valores de Nmin entre 25.5 y 27.9, NO3-N entre 4 y 4.8, y NH4+-N de 20.1 a 22.5. En el tratamiento N60, Nmin varía de 25 a 28.8, NO3-N de 3.3 a 5.5, y NH4+-N de 21.3 a 23.7. Finalmente, el tratamiento N80 muestra los valores más altos, con Nmin entre 28.5 y 37.1, NO3-N de 5.4 a 7.6, y NH4+-N de 22.9 a 29.7. En general, se observa un aumento en los valores de Nmin, NO3-N y NH4+-N a medida que aumenta la fertilización de N.

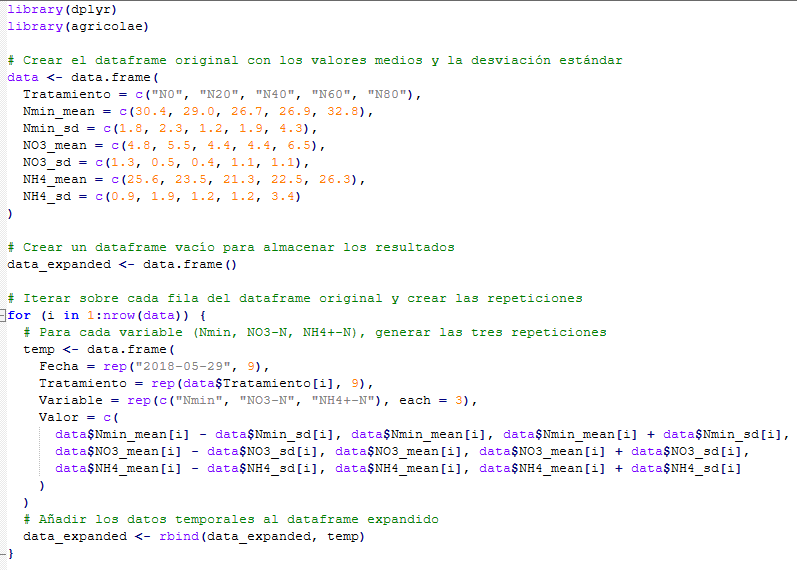
**Tabla 1** Datos organizados en bloque

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tratamiento** | **Emisores** | **N2O** | **Tratamiento** | **Emisores** | **N2O** |
| NH4+-N | N0 | 24.7 | Nmin | N80 | 28.5 |
| NH4+-N | N0 | 25.6 | Nmin | N80 | 32.8 |
| NH4+-N | N0 | 26.5 | Nmin | N80 | 37.1 |
| NH4+-N | N20 | 21.6 | NO3-N | N0 | 3.5 |
| NH4+-N | N20 | 23.5 | NO3-N | N0 | 4.8 |
| NH4+-N | N20 | 25.4 | NO3-N | N0 | 6.1 |
| NH4+-N | N40 | 20.1 | NO3-N | N20 | 5 |
| NH4+-N | N40 | 21.3 | NO3-N | N20 | 5.5 |
| NH4+-N | N40 | 22.5 | NO3-N | N20 | 6 |
| NH4+-N | N60 | 21.3 | NO3-N | N40 | 4 |
| NH4+-N | N60 | 22.5 | NO3-N | N40 | 4.4 |
| NH4+-N | N60 | 23.7 | NO3-N | N40 | 4.8 |
| NH4+-N | N80 | 22.9 | NO3-N | N60 | 3.3 |
| NH4+-N | N80 | 26.3 | NO3-N | N60 | 4.4 |
| NH4+-N | N80 | 29.7 | NO3-N | N60 | 5.5 |
| Nmin | N0 | 28.6 | NO3-N | N80 | 5.4 |
| Nmin | N0 | 30.4 | NO3-N | N80 | 6.5 |
| Nmin | N0 | 32.2 | NO3-N | N80 | 7.6 |
| Nmin | N20 | 26.7 |
| Nmin | N20 | 29 |
| Nmin | N20 | 31.3 |
| Nmin | N40 | 25.5 |
| Nmin | N40 | 26.7 |
| Nmin | N40 | 27.9 |
| Nmin | N60 | 25 |
| Nmin | N60 | 26.9 |
| Nmin | N60 | 28.8 |

# RESULTADOS Y DISCUSIONES

***5.1. Desarrollo de script en R***

El código en R está diseñado para preparar los datos para un análisis de un Diseño de Bloques Completos al Azar (DBCA). Primero, se cargan las bibliotecas necesarias dplyr y agricolae para el manejo de datos y análisis estadístico. Luego, se crea un dataframe con valores medios y desviaciones estándar para tres variables (Nmin, NO3-N y NH4+-N) en cinco tratamientos diferentes (N0, N20, N40, N60, N80). Posteriormente, el código genera un nuevo dataframe que expande los datos originales, creando tres repeticiones por tratamiento para cada variable. Cada repetición representa un valor mínimo (media - desviación estándar), un valor medio y un valor máximo (media + desviación estándar) para cada tratamiento y variable. Finalmente, el dataframe expandido se imprime, mostrando la estructura de datos necesaria para realizar un análisis DBCA con múltiples repeticiones por tratamiento y variable.



**Figura 4.** Llamado de librerías y de datos

Este código realiza un análisis de varianza (ANOVA) en los datos de emisiones de N₂O utilizando los diferentes **Emisores** como factor explicativo. La primera línea ajusta el modelo ANOVA con la función aov() para evaluar si las emisiones de N₂O varían significativamente entre los grupos de emisores.

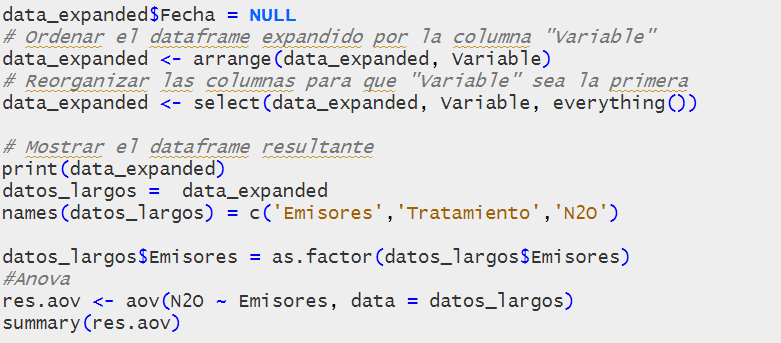
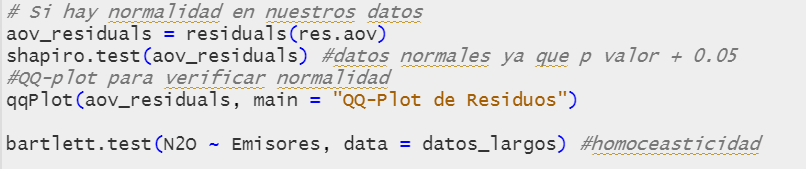


Figura 5 Código para crear los factores Tratamiento y Emisores para posterior test de ANOVA

Los resultados del ANOVA indican que existen diferencias altamente significativas en las emisiones de N₂O entre los diferentes grupos de emisores. Con un valor **F** de **400.3** y un valor **p < 2e-16**, mucho menor que el umbral común de 0.05, se rechaza la hipótesis nula de que las medias de los grupos son iguales, lo que sugiere que los emisores tienen un efecto significativo en las emisiones de N₂O. Los grados de libertad para los emisores son 2, y la suma de cuadrados es de **4783**, lo que refleja que una gran parte de la variación en los datos se debe a las diferencias entre emisores. Los residuos tienen una suma de cuadrados de **251** con 42 grados de libertad, lo que indica que la variabilidad dentro de los grupos es relativamente pequeña en comparación con la variabilidad entre los grupos.

El código realiza pruebas de supuestos sobre los residuos del modelo ANOVA previamente ajustado. Primero, se calculan los residuos del modelo res.aov y se almacenan en aov\_residuals. Luego, se aplica la prueba de Shapiro-Wilk (shapiro.test) para verificar si los residuos siguen una distribución normal. Posteriormente, se realizan dos pruebas de Bartlett (bartlett.test) para evaluar la homogeneidad de las varianzas (homocedasticidad). La primera prueba de Bartlett evalúa si las varianzas de "Valor" son homogéneas entre los diferentes niveles de la “Emisores”, mientras que la segunda prueba evalúa la homogeneidad de las varianzas del "Valor" entre los diferentes niveles del "Tratamiento". Estas pruebas son cruciales para validar los supuestos del ANOVA, que requieren normalidad de los residuos y homogeneidad de varianzas.

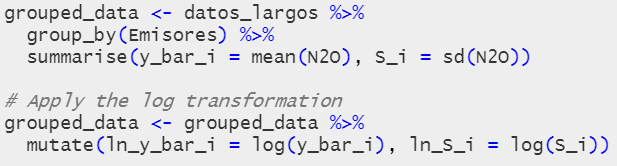


**Figura 6.** Líneas de código para aplicar el test de barttlet

Los resultados obtenidos muestran que los residuos del ANOVA se sometieron a la prueba de normalidad de Shapiro-Wilk, que arrojó un valor W de 0.9739 y un valor p de 0.3977. Dado que el valor p es mayor que 0.05, no hay evidencia suficiente para rechazar la hipótesis nula de normalidad, lo que indica que los residuos siguen una distribución normal, cumpliendo así con uno de los supuestos del ANOVA.

En cuanto a la homogeneidad de varianzas, la prueba de Bartlett para la “Emisores” mostró un valor p de 0.002271, lo que indica que las varianzas no son homogéneas (rechazo de la hipótesis nula de igualdad de varianzas). En resumen, los residuos son normales, pero existe una diferencia significativa en las varianzas entre los niveles de la "Emisores", mientras que las varianzas entre los tratamientos son homogéneas.

El objetivo del siguiente código es agrupar los datos por el tipo de **Emisores** y calcular dos estadísticas clave para cada grupo: la media (​) y la desviación estándar (​) de las emisiones de N₂O. Luego, se aplica una transformación logarítmica a ambas estadísticas (media y desviación estándar), generando nuevas variables log-transformadas: ​ y Esta transformación es útil para estabilizar la varianza, linealizar relaciones no lineales, y puede ser particularmente beneficiosa para datos que no siguen una distribución normal, facilitando el análisis posterior de los patrones en los diferentes grupos de emisores.



El objetivo de este código es representar gráficamente la relación entre el logaritmo de la desviación estándar y el logaritmo de la media ​ para diferentes grupos de emisores de N₂O. Primero, ajusta un modelo de regresión lineal a estos datos, generando una ecuación que describe esta relación, y luego agrega la línea de regresión al gráfico. Además, el código calcula y muestra la ecuación de la regresión en el gráfico, proporcionando una representación visual de la relación entre la variabilidad y el promedio de las emisiones log-transformadas para los diferentes grupos.



Se obtuvo y por lo que la transformación más próxima es la de raíz cuadrada.

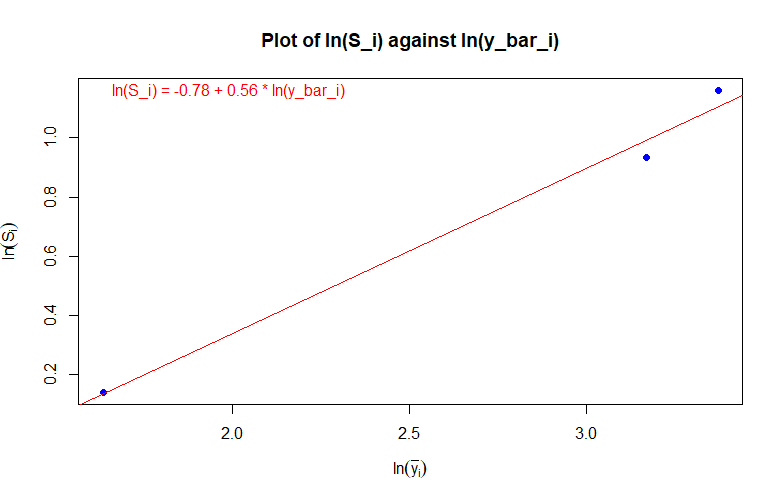
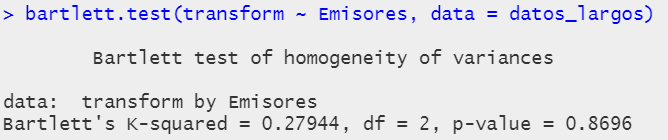


Figura 7 ajuste de modelo lineal a ​ vs

Para evaluar la homogeneidad de varianzas en los diferentes grupos de emisores de N₂O, se aplicó la prueba de Bartlett. Inicialmente, se observó que los datos sin transformar no cumplían con el supuesto de homocedasticidad, ya que el resultado de la prueba fue significativo (**K-squared = 12.175**, **p-value = 0.0023**), lo que indica que las varianzas entre los grupos de emisores eran significativamente diferentes. Sin embargo, al aplicar una transformación a los datos, el test de Bartlett mostró un resultado no significativo (**K-squared = 0.27944**, **p-value = 0.8696**), confirmando que las varianzas de los grupos se igualaron tras la transformación. Por lo tanto, se concluye que los datos transformados cumplen con el supuesto de homogeneidad de varianzas, permitiendo la realización de un ANOVA clásico sobre los datos transformados para comparar los emisores de N₂O.



Los resultados del ANOVA indican que hay una diferencia altamente significativa en las emisiones de N₂O transformadas entre los distintos grupos de emisores. El valor F es muy elevado (F = 597.3) y el valor p es extremadamente pequeño (p < 2e-16), lo que sugiere que las diferencias entre los grupos de emisores no son debidas al azar. Esto significa que los diferentes emisores (NH4+-N, NO3-N y Nmin) tienen un impacto significativo sobre las emisiones transformadas de N₂O. Los residuos tienen una suma de cuadrados pequeña (3.00), lo que indica que el modelo ANOVA explica una gran parte de la variabilidad en los datos. En resumen, los emisores influyen de manera significativa en las emisiones de N₂O en el conjunto de datos transformados.



# CONCLUSIONES Y RECOMENDACIONES

En el análisis inicial de la varianza (ANOVA) entre los emisores de N₂O, se observó que las emisiones difieren significativamente entre los distintos grupos de emisores (NH4+-N, NO3-N y Nmin). El valor de **F** obtenido fue de **400.3**, con un valor p extremadamente pequeño (**p < 2e-16**), lo que confirma que las diferencias no son producto del azar y que los emisores tienen un impacto significativo sobre las emisiones de N₂O. Sin embargo, al revisar los supuestos del ANOVA, la prueba de Bartlett mostró que las varianzas entre los grupos no eran homogéneas (**p = 0.0023**), lo que viola uno de los supuestos clave del análisis. Además, la prueba de Durbin-Watson indicó autocorrelación en los residuos del modelo (**p = 0.00133**), lo que sugiere que las observaciones no son completamente independientes.

Dado que no se cumplían los supuestos de homogeneidad de varianzas ni de independencia, se aplicó una transformación de raíz cuadrada a los datos de emisiones de N₂O. Al volver a realizar el ANOVA con los datos transformados, se observó nuevamente una diferencia significativa entre los emisores (**F = 597.3**, **p < 2e-16**). La transformación de los datos mejoró notablemente el cumplimiento de los supuestos del análisis, ya que la prueba de Bartlett no rechazó la hipótesis de homogeneidad de varianzas (**p = 0.8696**), lo que indica que las varianzas de los grupos son ahora homogéneas. Este resultado valida el uso del ANOVA clásico para los datos transformados.

A pesar de la mejora en la homocedasticidad, la prueba de Durbin-Watson en los datos transformados aún indicó una ligera autocorrelación en los residuos (**p = 0.00807**), aunque menor que en el análisis inicial. La presencia de autocorrelación sugiere que las observaciones no son completamente independientes entre sí, lo que puede afectar ligeramente la interpretación de los resultados. No obstante, el grado de autocorrelación residual es menor en los datos transformados, lo que hace que los resultados del ANOVA sean más robustos que en el análisis sin transformación.

Finalmente, se exploró la relación entre la desviación estándar y la media de las emisiones de N₂O mediante la transformación logarítmica de ambas variables. El análisis de regresión lineal entre y el logaritmo de la media ​ mostró una relación positiva con una pendiente de **0.5574**, lo que sugiere que las variaciones en las emisiones de N₂O aumentan a medida que lo hacen las medias de los emisores. Esta relación lineal es consistente con la idea de que, a mayores niveles de emisiones, también se presenta mayor variabilidad en los datos, lo que subraya la importancia de la transformación de datos para estabilizar esta relación y mejorar la validez del análisis estadístico.

# REFERENCIAS BIBLIOGRÁFICAS

Atkinson, A. C., Riani, M., & Corbellini, A. (2021). The Box-Cox transformation: Review and extensions. *Statistical Science, 36*(2), 239-255. <https://doi.org/10.1214/20-STS778>

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>

Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression*. Chapman & Hall.

Chung, M., Haber, E. (2012). Experimental Design for Biological Systems. SIAM Journal on Control and Optimization, 50, 471–489. <https://doi.org/10.1137/100791063>

Durbin, J., Watson, G. S. (1950). Testing for Serial Correlation in Least Squares Regression: I. Biometrika, 37(3–4), 409–428. <https://doi.org/https://doi.org/10.2307/2332391>

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Hayter, A. J. (1984). March, 1984 A Proof of the Conjecture that the Tukey-Kramer Multiple Comparisons Procedure is Conservative. Ann. Statist, 12(1), 61–75. <https://doi.org/10.1214/aos/1176346392>

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika, 36*(1/2), 149-176.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). SAGE Publications.

Montgomery, D. C. (2013). Design and Analysis of Experiments (A. S. University (ed.); Eighth Edi). John Wiley & Sons, Inc.

Mynbaev, K. T., & Darkenbayeva, G. S. (2019). Analyzing variance in central limit theorems. *Kazakh Mathematical Journal, 19*(3), 30-39.

<https://mpra.ub.uni-muenchen.de/101685/>

Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, 15(12), 1-9. ISSN 1531-7714

Razali, N.M., Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics, 2(1), 21–33.

Smit, H. P. J., Reinsch, T., Swanepoel, P. A., Kluß, C., & Taube, F. (2020). Grazing under irrigation affects N₂O-emissions substantially in South Africa. *Atmosphere*, 11(9), 925. <https://doi.org/10.3390/atmos11090925>

Snedecor, G.W., Cochran, W. G. (1989). Statistical Methods (IOWA STATE UNIVERSITY PRESS (ed.); 8th Editio). Library or Congress Cataloging-in-Publication Data.

# ANEXOS

**Anexo 1.** Script de los datos analizados

# Cargamos las librerías necesarias

library**(**tidyr**)**

library**(**dplyr**)**

library**(**agricolae**)**

# Cargar la librería ggplot2

library**(**ggplot2**)**

library**(**lmtest**)**

library**(**car**)** # Para la verificación de suposiciones

# Crear el dataframe original con los valores medios y la desviación estándar

data **<-** data.frame**(**

Tratamiento **=** c**(**"N0", "N20", "N40", "N60", "N80"**)**,

Nmin\_mean **=** c**(**30.4, 29.0, 26.7, 26.9, 32.8**)**,

Nmin\_sd **=** c**(**1.8, 2.3, 1.2, 1.9, 4.3**)**,

NO3\_mean **=** c**(**4.8, 5.5, 4.4, 4.4, 6.5**)**,

NO3\_sd **=** c**(**1.3, 0.5, 0.4, 1.1, 1.1**)**,

NH4\_mean **=** c**(**25.6, 23.5, 21.3, 22.5, 26.3**)**,

NH4\_sd **=** c**(**0.9, 1.9, 1.2, 1.2, 3.4**)**

**)**

# Crear un dataframe vacío para almacenar los resultados

data\_expanded **<-** data.frame**()**

# Iterar sobre cada fila del dataframe original y crear las repeticiones

**for** **(**i **in** 1**:**nrow**(**data**))** **{**

# Para cada variable (Nmin, NO3-N, NH4+-N), generar las tres repeticiones

temp **<-** data.frame**(**

Fecha **=** rep**(**"2018-05-29", 9**)**,

Tratamiento **=** rep**(**data**$**Tratamiento**[**i**]**, 9**)**,

Variable **=** rep**(**c**(**"Nmin", "NO3-N", "NH4+-N"**)**, each **=** 3**)**,

Valor **=** c**(**

data**$**Nmin\_mean**[**i**]** **-** data**$**Nmin\_sd**[**i**]**, data**$**Nmin\_mean**[**i**]**, data**$**Nmin\_mean**[**i**]** **+** data**$**Nmin\_sd**[**i**]**,

data**$**NO3\_mean**[**i**]** **-** data**$**NO3\_sd**[**i**]**, data**$**NO3\_mean**[**i**]**, data**$**NO3\_mean**[**i**]** **+** data**$**NO3\_sd**[**i**]**,

data**$**NH4\_mean**[**i**]** **-** data**$**NH4\_sd**[**i**]**, data**$**NH4\_mean**[**i**]**, data**$**NH4\_mean**[**i**]** **+** data**$**NH4\_sd**[**i**]**

**)**

**)**

# Añadir los datos temporales al dataframe expandido

data\_expanded **<-** rbind**(**data\_expanded, temp**)**

**}**

data\_expanded**$**Fecha **=** **NULL**

# Ordenar el dataframe expandido por la columna "Variable"

data\_expanded **<-** arrange**(**data\_expanded, Variable**)**

# Reorganizar las columnas para que "Variable" sea la primera

data\_expanded **<-** select**(**data\_expanded, Variable, everything**())**

# Mostrar el dataframe resultante

print**(**data\_expanded**)**

datos\_largos **=** data\_expanded

names**(**datos\_largos**)** **=** c**(**'Emisores','Tratamiento','N2O'**)**

datos\_largos**$**Emisores **=** as.factor**(**datos\_largos**$**Emisores**)**

#Anova

res.aov **<-** aov**(**N2O **~** Emisores, data **=** datos\_largos**)**

summary**(**res.aov**)**

# Si hay normalidad en nuestros datos

aov\_residuals **=** residuals**(**res.aov**)**

shapiro.test**(**aov\_residuals**)** #datos normales ya que p valor + 0.05

#QQ-plot para verificar normalidad

qqPlot**(**aov\_residuals, main **=** "QQ-Plot de Residuos"**)**

bartlett.test**(**N2O **~** Emisores, data **=** datos\_largos**)** #homoceasticidad

#Durbin-Watson test

lmtest**::**dwtest**(**res.aov**)** # no hay autocorrelacion

tukeyMetodo2 **=** HSD.test**(**res.aov,"Emisores", alpha **=**0.05, group**=**T**)**

plot**(**tukeyMetodo2,variation**=**"IQR"**)**

############################################################

################Transformacion de datos#####################

############################################################

################################

grouped\_data **<-** datos\_largos %>%

group\_by**(**Emisores**)** %>%

summarise**(**y\_bar\_i **=** mean**(**N2O**)**, S\_i **=** sd**(**N2O**))**

# Apply the log transformation

grouped\_data **<-** grouped\_data %>%

mutate**(**ln\_y\_bar\_i **=** log**(**y\_bar\_i**)**, ln\_S\_i **=** log**(**S\_i**))**

# Plot ln(S\_i) against ln(y\_bar\_i)

plot**(**grouped\_data**$**ln\_y\_bar\_i, grouped\_data**$**ln\_S\_i,

xlab **=** expression**(**ln**(**bar**(**y**)[**i**]))**, ylab **=** expression**(**ln**(**S**[**i**]))**,

main **=** "Plot of ln(S\_i) against ln(y\_bar\_i)",

pch **=** 19, col **=** "blue"**)**

# Adding a linear regression line

model **<-** lm**(**ln\_S\_i **~** ln\_y\_bar\_i, data **=** grouped\_data**)**

abline**(**model, col **=** "red"**)**

print**(**model**)**

# Extract coefficients

intercept **<-** round**(**coef**(**model**)[**1**]**, 2**)** # intercept

slope **<-** round**(**coef**(**model**)[**2**]**, 2**)** # slope

# Create the equation text

equation **<-** paste0**(**"ln(S\_i) = ", intercept, " + ", slope, " \* ln(y\_bar\_i)"**)**

# Add the equation to the plot

text**(**x **=** min**(**grouped\_data**$**ln\_y\_bar\_i**)**, y **=** max**(**grouped\_data**$**ln\_S\_i**)**,

labels **=** equation, pos **=** 4, col **=** "red"**)**

# Calculate alpha and lambda

alpha **<-** model**$**coefficients**[**2**]**

landa **<-** 1 **-** alpha

landa

datos\_largos**$**transform **=** sqrt**(**datos\_largos**$**N2O**)**

datos\_largos**$**Emisores **=** as.factor**(**datos\_largos**$**Emisores**)**

res.aov2 **<-** aov**(**transform **~** Emisores, data **=** datos\_largos**)**

summary**(**res.aov2**)**

bartlett.test**(**transform **~** Emisores, data **=** datos\_largos**)** #homoceasticidad

#Durbin-Watson test

lmtest**::**dwtest**(**res.aov2**)** # no hay autocorrelacion

tukeyMetodo2 **=** HSD.test**(**res.aov2,"Emisores", alpha **=**0.05, group**=**T**)**

plot**(**tukeyMetodo2,variation**=**"IQR"**)**

# Crear el histograma de N2O, coloreando por el tipo de Emisores

ggplot**(**datos\_largos, aes**(**x **=** N2O, fill **=** Emisores**))** **+**

geom\_histogram**(**binwidth **=** 0.5, color **=** "black", alpha **=** 0.7**)** **+** # Histograma con contorno negro y transparencia

labs**(**title **=** "Histograma de N2O por Emisores", x **=** "N2O", y **=** "Frecuencia"**)** **+** # Etiquetas de ejes y título

theme\_minimal**()** **+** # Tema limpio

scale\_fill\_brewer**(**palette **=** "Set3"**)** # Paleta de colores para diferenciar emisores

# Crear el histograma de N2O, coloreando por el tipo de Emisores

ggplot**(**datos\_largos, aes**(**x **=** transform, fill **=** Emisores**))** **+**

geom\_histogram**(**binwidth **=** 0.5, color **=** "black", alpha **=** 0.7**)** **+** # Histograma con contorno negro y transparencia

labs**(**title **=** "Histograma de transform por Emisores", x **=** "N2O transformado", y **=** "Frecuencia"**)** **+** # Etiquetas de ejes y título

theme\_minimal**()** **+** # Tema limpio

scale\_fill\_brewer**(**palette **=** "Set3"**)** # Paleta de colores para diferenciar emisores