

NOTA DE INVESTIGACION

APLICACIÓN DE TRANSFORMACIONES PARA EL CUMPLIMIENTO DE LOS SUPUESTOS DE NORMALIDAD Y HOMOCEDASTICIDAD, A CONCENTRACIONES FOLIARES DE N, P Y K EN MANDARINO

Avanza, María ⁽¹⁾ ; Mazza, Silvia ⁽¹⁾ ; Martínez, Gloria ⁽¹⁾ ; Giménez, Laura ⁽¹⁾⁽²⁾

⁽¹⁾ Facultad de Ciencias Agrarias- UNNE⁽²⁾ INTA- EEA Sombrerito.

E-mail: mavanza@agr.unne.edu.ar

RESUMEN

El objetivo del presente trabajo consistió en analizar la distribución que presentan las concentraciones foliares de N (g.kg^{-1}), P (g.kg^{-1}) y K (g.kg^{-1}) en mandarino (*Citrus reticulata*, Blanco), y evaluar el efecto de diferentes transformaciones aritméticas sobre su normalidad y homocedasticidad. Las variables en estudio fueron determinadas en 12 muestreos de ramas fructíferas realizados durante las estaciones de verano, otoño e invierno, en lotes comerciales ubicados en el Departamento de Monte Caseros, provincia de Corrientes. Para evaluar la normalidad se utilizó el test de Shapiro-Wilk y el test de Shapiro-Wilk modificado por Mahibbur y Govindarajulu (1997), complementando el análisis con herramientas descriptivas, entre ellas gráfico de frecuencia y de Box Plot. La homocedasticidad se analizó por medio de los tests de Bartlett y Levene. Se trabajó sobre las variables originales y las familias de transformaciones: logaritmo natural, raíz cuadrada y potencia de Box y Cox. Las transformaciones de potencia de Box y Cox son las que logran establecer la normalidad para N y K, en cambio para P no se consigue normalidad con ninguna de las transformaciones probadas. Según el test de Bartlett, las transformaciones logarítmicas y raíz cuadrada reducen la heterocedasticidad pero no logran homogeneizarlas en todos los casos. En cambio, las transformaciones de potencia de Box y Cox son las más efectivas para lograr la homocedasticidad en estas variables. Los valores obtenidos con la prueba de Levene, en todos los casos indican el logro de la homocedasticidad.

Palabras claves: *Citrus reticulata*, mandarino, concentraciones foliares de N, concentraciones foliares de P, concentraciones foliares de K.

SUMMARY

The objective of the this work consisted in the analysis of the distribution of N (g.kg^{-1}), P (g.kg^{-1}) and K (g.kg^{-1}) leaves concentration in mandarin orange (*Citrus reticulata*, Blanco) and to evaluate the effect of the different arithmetic transformations on normality and homocedasticity. The variables studied were determined on 12 samples of fruitful branches, during summer, autumn and winter in comercial lots placed at Department of Monte Caseros, Province of Corrientes. Shapiro- Wilk and the Shapiro- Wilk modified by Mahibbur and Govindarajulu (1997) tests, were used to evaluate normality, adding for the analysis descriptive, frequency graphics and the Box Plot. Variance homogeneity was analyzed by Bartlett and Levene test. Statistical analysis on the original variables and the transformation families: natural logarithm, square root and power of Box and Cox were performed. Box and Cox power transformations established normality for N and K; for P foliar concentration. No one of the tested transformations meet normality. According with Bartlett test, logarithmic and square root transformation not always reduce the heterocedasticity. Although Box and Cox power transformations are most effective to achieve the homocedasticity. According with Levene test, all transformations achieve homogeneity.

Key words: *Citrus reticulata*, mandarin, orange, N leaves concentration, P leaves concentration, K leaves concentration.

INTRODUCCIÓN

El estudio del contenido de los elementos minerales, que entran en la composición de los tejidos foliares de los citrus, refleja en gran medida el estado nutritivo de los árboles en una situación dada. Los elementos minerales fundamentales de la nutrición de los citrus son nitrógeno, fósforo y potasio, los que se completan con calcio y magnesio. En la mayoría de los casos, solo los tres primeros son objeto de aportes regulares en las prácticas citrícolas (Loussert, 1992).

El diagnóstico foliar es una técnica utilizada en plantaciones citrícolas del mundo entero, para reconocer y/o prevenir las carencias o excesos de nutrientes minerales, antes de que éstos se manifiesten visualmente en los árboles, lo que permite una mayor precisión en los programas de fertilización (Fabiani, 1996).

En los trabajos de Investigación Citrícola, los contenidos foliares de nutrientes constituyen variables respuesta utilizadas frecuentemente y por lo tanto pasibles de análisis estadísticos bajo modelos lineales.

Los modelos de análisis estadístico, como el análisis de varianza, son sensibles a las propiedades estadísticas de los términos de error aleatorio del modelo lineal. Los supuestos del análisis de varianza implican errores independientes, normalmente distribuidos y con varianzas homogéneas para todas las observaciones, condiciones que muchas veces no se cumplen (Infostat, 2002).

Las distribuciones de los contenidos foliares de los nutrientes presentan fallas en algunos de los supuestos del análisis de variancia, los más graves son el apartamiento de la distribución normal y la falta de homogeneidad entre las variancias de los tratamientos (heterocedasticidad).

Ante situaciones en que existen fuertes dudas sobre el cumplimiento de los supuestos, se pueden seguir varios caminos: si las desviaciones en las hipótesis de cumplimiento no son muy grandes, de igual manera se aplica la técnica teniendo en cuenta las faltas en que se incurre y que la técnica en ese caso es aproximada; se buscan técnicas alternativas que no exijan los supuestos no cumplidos, como técnicas no paramétricas, Mínimos Cuadrados Ponderados o los Modelos Lineales Generalizados. Se aplican transformaciones aritméticas a los datos que logren solucionar las desviaciones de los supuestos. Además, hay situaciones en que puede requerirse simplificar el modelo, reduciendo el orden o eliminando interacciones, a los fines de una mejor interpretación (Abbiati, 1990; Steel y Torrie, 1992).

La transformación aritmética de datos, es una técnica ampliamente utilizada previa al análisis estadístico. Las razones básicas para transformar

la información son: hacer el análisis lo más simple posible y/o encontrar una métrica en la cual los supuestos teóricos necesarios para el análisis se satisfagan. Transformar la variable respuesta, de manera que conduzca al cumplimiento simultáneo de los supuestos y si es necesario simplifique el modelo, permitiendo el empleo del análisis de variancia tradicional, es una alternativa muy utilizada (Abbiati, Recchioni, 1990; Bartlett, 1947; Steel y Torrie, 1992).

Hoyle (1971), indica que son 19 las transformaciones de uso corriente. En general las más utilizadas son las transformaciones pertenecientes a las familias raíz cuadrada, logarítmica y angulares (Hoyle, 1971; Steel y Torrie, 1992).

Box y Cox (1964), estudian la familia de transformaciones de potencia, cambiando el valor λ , evaluando en forma sencilla el efecto de las mismas sobre los supuestos de análisis de varianza.

Mazza, *et al.* (2000), evalúan el efecto de varias familias de transformaciones sobre la normalidad y la homocedasticidad de las concentraciones foliares de Zn y K en naranjo (*Citrus sinensis*, L.) var. Valencia, encontrando mejor comportamiento con las transformaciones logarítmicas en el caso del K y las de raíz cuadrada para el Zn.

Los objetivos del presente trabajo fueron determinar la distribución que presentan las variables en estudio y evaluar el efecto de diferentes transformaciones aritméticas sobre su normalidad y homocedasticidad.

MATERIALES Y MÉTODOS

La información utilizada proviene de 12 muestreos de ramas fructíferas de mandarina (*Citrus reticulata*, Blanco), realizados durante las estaciones de verano, otoño e invierno de los años 1997 al 1999, en lotes comerciales ubicados en el Departamento de Monte Caseros, provincia de Corrientes.

Las variables analizadas fueron concentraciones foliares de N(g.Kg⁻¹), P(g.Kg⁻¹) y K(g.Kg⁻¹) y las resultantes de cada una de las familias de las transformaciones: logarítmicas, raíces cuadradas y de potencia de Box y Cox (Abbiati y Recchioni, 1990).

La familia de transformaciones de potencias propuestas por Box y Cox se expresa: $y(\lambda) = (y^{\lambda} - 1) / \lambda$. Esta familia continua, depende del parámetro λ y la variable respuesta debe ser positiva. Para realizar el cálculo de λ , se comienza utilizando un intervalo de confianza de [-2, 2] que posibilita hallar una transformación con mayor sentido práctico (Abbiati y Recchioni, 1990).

Para evaluar la normalidad se utilizó el test de Wilk-Shapiro y el de Wilk-Shapiro modificado por Mahibbur y Govindarajulu (1997), complementando el análisis con herramientas descriptivas, entre ellas gráfico de frecuencia y de Box Plot (Infostat, 2002).

Tanto los histogramas como los diagramas de caja (Box Plot) tienen por objeto representar una distribución de frecuencias. La diferencia está en que en los histogramas se puede observar claramente la asimetría de una distribución, pero es difícil, a partir de ellos, visualizar la forma de la distribución de frecuencias de cada grupo de observaciones. El gráfico de cajas (Box Plot), tiene por objeto reflejar mejor la forma de estas distribuciones dando en un mismo elemento gráfico información acerca de la mediana, la media, cuantiles 0.05, 0.25, 0.75 y 0.95 y mostrando presencia, si los hubiere, de valores extremos (Di Rienzo *et al.*, 2000).

El test de Wilk- Shapiro calcula el estadístico W, que prueba la normalidad de la distribución.

El test de Wilk- Shapiro modificado por Mahibbur y Govindarajulu (1997), utiliza los residuos como variable de análisis y con el estadístico W somete a prueba la hipótesis distribución normal de esos residuos.

La homogeneidad de variancias (homocedasticidad) se analizó por medio de los tests de Bartlett y Levene.

La prueba de Bartlett es una modificación de la prueba de razón de verosimilitud de Neyman-Pearson, donde el estadístico final de significación es X^2 , el cual se compara con un valor crítico de χ^2 (Steel y Torrie, 1992). Por lo tanto esta prueba se basa en que, en la hipótesis nula de igualdad de variancias y poblaciones normales, un estadístico calculado a partir de las variancias muestrales y el cuadrado medio del error sigue una distribución chi-cuadrado (www.hrc.es/biost./Anova).

El test de Levene consiste en realizar una ANOVA usando como variable dependiente el valor absoluto de los residuos. Si el valor p del factor tratamiento de este ANOVA es menor al valor de significación nominal se rechaza la hipótesis de varianzas homogéneas, caso contrario el supuesto de igualdad de varianzas puede ser sostenido (O'Neil y Matheus, 2002).

Estos análisis se realizaron con los paquetes estadísticos Infostat (Infostat, 2002) de la propiedad de la Sec. Gral. de Ciencia y Técnica UNNE, y SAS (SAS, 2000), disponible a través del convenio INTA-UNNE.

RESULTADOS Y DISCUSIÓN

En un análisis descriptivo se determinaron las medidas de posición y dispersión de las variables bajo estudio. La medida de tendencia central o

de posición más común es la media aritmética, con valores para N= 2.03, P=0.14 y K=0.63 y para las variables transformadas por la potencia de Box y Cox son N=10.93, P= 10.90 y K= 5.88; donde estos valores dan una idea bastante aproximada de donde se encuentra localizado o centrado el histograma.

El coeficiente de variación (CV) es una medida relativa de variación, en contraste con el desvío estándar y se define como el cociente entre la desviación estándar y la media aritmética, expresada en porcentaje.

El CV arroja valores para las variables N= 18.10, P= 26.87 y K= 30.81 y para las variables transformadas por la potencia de Box y Cox N=0.14, P=2.51 y K=1.22. Esto indica que existe una gran variabilidad en el conjunto de los datos de las variables originales y para las variables transformadas esa variabilidad disminuye.

Los histogramas de las variables originales (Figura 1) son típicos de muchas distribuciones de frecuencias obtenidas de datos que se encuentran en la naturaleza, se puede observar que los tres casos (N, P y K) presentan distribución asimétrica hacia la derecha, donde esta asimetría indica la falta de simetría con respecto al eje vertical que lo constituye la media aritmética. En los histogramas de las variables transformadas por la potencia de Box y Cox (Figura 1), se observa una asimetría menos marcada que en el caso de las distribuciones originales. La distribución es asimétrica con valores muy extremos N=3.00, P=0.25 y K=1.25, a la derecha de la mediana; y para las variables transformadas por potencia de Box y Cox la asimetría es menos marcada y con valores extremos en el caso de P.

Los resultados del Test de Wilk-Shapiro (Tabla 1), indican que las distribuciones de N, P y K se alejan de la normal, y que las transformaciones logarítmicas consiguen una aproximación a la distribución normal, pero las de potencia son las que realmente logran establecer la normalidad para N y K. En relación al test de Wilk-Shapiro modificado (Tabla 1), que consiste en seleccionar los residuos como variable del test, los resultados indican que, tanto las transformaciones logarítmicas como las de potencia de Box y Cox, consiguen la distribución normal de las variables N y K. Para la variable P no se logra establecer la normalidad en ninguno de los casos.

Los valores obtenidos para el test de Bartlett (Tabla 2), muestran que para las concentraciones foliares del N y P en las cuales los errores son heterocedásticos, las transformaciones logarítmicas y de raíz cuadrada reducen la heterogeneidad de variancias pero no logran homogeneizarlas. En cambio, las transformaciones de potencia de Box y Cox son las más efectivas para lograr la homocedasticidad

en estas dos variables. En el caso del contenido foliar de K los errores son homocedásticos. Cabe resaltar que este test es muy sensible a la falta de normalidad de los datos, por lo tanto puede indicar la no normalidad en vez de heterocedasticidad.

Los valores obtenidos con la prueba de Levene, en todos los casos indican el logro de la homocedasticidad, debido a que el valor p es mayor al valor de significación nominal ($\alpha=0.05$).

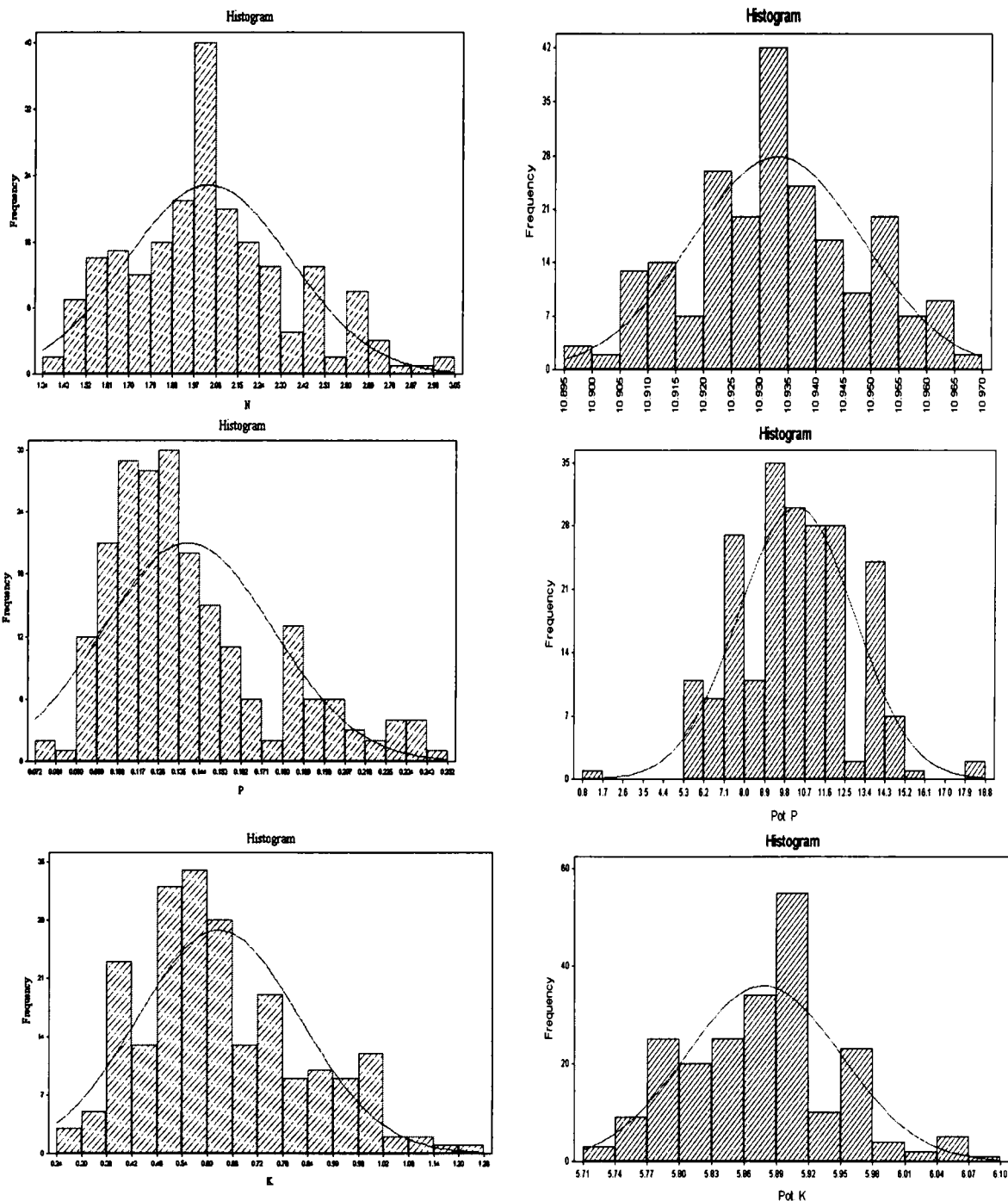


Figura 1: Histogramas que representan las distribuciones del N, P, K y sus respectivas transformaciones de potencia de Box y Cox.

Tabla 1: Evaluación de la Normalidad por el Test de Wilk- Shapiro y el Test de Wilk- Shapiro modificado por Mahibbur y Govindarajulu (1997).

Variables	W	valor P	W (modificado)	valor P
N	0.97	0.0007	0.97	0.0017
P	0.92	<0.0001	0.93	<0.0001
K	0.95	<0.0001	0.94	<0.0001
Logaritmo N	0.99	0.0473	0.98	0.1610
Logaritmo P	0.97	0.0002	0.95	<0.0001
Logaritmo K	0.98	0.0279	0.98	0.0793
Raíz cuadrada N	0.98	0.0128	0.98	0.0231
Raíz cuadrada P	0.95	<0.0001	0.94	<0.0001
Raíz cuadrada K	0.98	0.0021	0.96	<0.0001
Pot. Box y Cox de N = -0.10	0.99	0.0496	0.98	0.2001
Pot. Box y Cox de P = -1.10	0.98	0.0068	0.93	<0.0001
Pot. Box y Cox de K = -0.21	0.98	0.0121	0.99	0.3079

Tabla 2: Evaluación de la homogeneidad de variancias entre los diferentes muestreos. Test de Bartlett y Levene.

Variables	Bartlett		Levene	
	Chi-SQ	valor P	F	valor P
N	6.78	0.0338	1.64	0.0889
P	20.72	0.0000	1.70	0.0742
K	0.80	0.6692	1.26	0.2487
Logaritmo N	0.47	0.7914	0.91	0.5355
Logaritmo P	6.45	0.0397	1.51	0.1293
Logaritmo K	0.11	0.9482	1.17	0.3102
Raíz cuadrada N	2.48	0.2890	1.19	0.2971
Raíz cuadrada P	12.42	0.0020	1.46	0.1490
Raíz cuadrada K	0.33	0.8459	1.01	0.4422
Pot. Box y Cox de N = -0.10	0.46	0.7964	0.88	0.5638
Pot. Box y Cox de P = -1.10	0.92	0.6320	2.18	0.0169
Pot. Box y Cox de K = -0.21	0.64	0.7256	1.35	0.1979

CONCLUSIONES

Las concentraciones foliares de N, P y K en mandarino no poseen una distribución que permita su análisis estadístico bajo modelos lineales, por lo que se recomienda, con el objeto de no afectar el nivel de significancia ni la sensibilidad de F respecto a la hipótesis nula en el ANOVA, aplicar previamente las transformaciones logarítmicas o de potencia de Box y Cox con valores del parámetro λ negativos y cercanos a cero.

BIBLIOGRAFÍA

- Abbiati, N; Recchioni, L. 1990. Transformaciones de potencias en modelos lineales. Comunicaciones en estadística aplicada a Ciencias Agropecuarias. INTA. Serie Divulgación N° 16: 1-13.
- Bartlett, M.; 1947. The use of transformations. *Biometrics*, (3): 39-52.
- Biblioteca Hospital Ramón y Cajal (Barcelona). Material docente de la unidad de Bioestadística Clínica [en línea]. Última actualización : 17 de mayo de 2002. Apuntes de Cursos de la Unidad. [Consulta: 12/07/02].
- Box, G.E.P and Cox, R.H. 1964. An analysis of transformations. *J.R. Statist. Soc. B.* (26): 211-252.
- Di Rienzo, J; Casanoves, F; Gonzales, L.; Tablado, E.; Díaz, M.; Robledo, C.; Balzarini, M. Estadística para las ciencias agropecuarias. 3er edición. FCA-UNC: 27-30.
- Fabiani, A.; Mika, R.; Larocca, L.; Anderson, C. 1996. Manual para productores de naranja y mandarina de la región del Río Uruguay. PRODIP-INTA: 96-99.
- Hoyle, M.H. 1971. Transformations- An Introduction and Bibliography. This review is part of the author's Ph.D. Thesis at the University of London.

- INFOSTAT (2002). Infostat versión 1.1- Grupo Infostat, FCA, Universidad Nacional de Córdoba, Argentina.
- INFOSTAT (2002). Infostat versión 1.1 Manual del usuario. Grupo Infostat, FCA. Universidad Nacional de Córdoba. Primera Edición. Ed. Brujas: 61-90.
- Loussert, R. 1992. Los agrios. Castelló- 37. Madrid: 189-191.
- Mahibbur, RM. Y Govindarajulu, Z. 1997. A modification of the test of Shapiro and Wilks for normality. *Journal of Applied Statistics*, 24(2): 219-235.
- Mazza, S.; Giménez, L.; Schroeder, J.; Martínez, G.; Rodríguez, V.; 2000. Concentraciones foliares de Zn y K. Efecto de transformaciones sobre supuestos de normalidad y homogeneidad de variancias. V Reunión Científica del Grupo Argentino de Biometría. Vaquerías- Córdoba.
- O Neil, E. And Mathews, L. 2002. Levene tests of homogeneity of variance for general block and treatment designs. *Biometrics*. (58) : 216-224.
- SAS Institute Inc, (2000). *Statistical Analysis Systems Version Eight*. Site. 14759001.
- Steel, R.G.D. y Torrie, J.H. 1992. *Bioestadística: Principios y procedimientos*. Mc. Graw Hill. México: 132-164; 458-461.