

# ¿Cómo proceder ante el incumplimiento de las premisas de los métodos paramétricos? o ¿cómo trabajar con variables biológicas no normales?

## How do we proceed to violations of parametric methods assumptions? or how to work with non-normal biological variables?

Leneidy Pérez Pelea\*

### RESUMEN

En las investigaciones biológicas se obtienen con frecuencia, datos experimentales que no cumplen con las premisas establecidas para poder aplicar un método estadístico paramétrico, como la normalidad y homocedasticidad. De manera general, cuando se incumple alguna o todas las premisas, los investigadores suelen aplicar una transformación de escala a los datos. Con el empleo de las transformaciones no siempre se logra el cumplimiento de las premisas de los análisis paramétricos. Se ha demostrado que su empleo logra mejorar el ajuste de los datos, pero puede complicar la interpretación de los resultados debido al cambio de escala de las variables. Otra alternativa consiste en emplear un método no paramétrico, que no requiere el cumplimiento de supuestos, pero si de un buen diseño del experimento, un tamaño de muestra adecuado y una correcta aleatorización de los tratamientos. Los métodos no paramétricos tienen menor potencia pues se incrementa la probabilidad de no detectar diferencias que si existen entre los tratamientos comparados. La probabilidad de tomar una decisión incorrecta se incrementa aún más si el tamaño de la muestra es pequeño y no se emplea el mismo número de réplicas en cada tratamiento. Existen otros análisis más complejos que requieren de métodos de cálculo más intensivo, como los métodos de remuestreo, aleatorización, permutaciones, bayesianos, el meta-análisis y los modelos lineales generalizados. Para poder interpretar los resultados obtenidos con estos análisis se requiere de un cambio en la mentalidad de los investigadores acostumbrados a trabajar con las herramientas de la estadística frecuentista.

**Palabras clave:** Transformaciones de escala, Métodos no paramétricos, Modelos Lineales Generalizados, Modelos nulos

### ABSTRACT

In biological research we frequently obtain experimental data which do not comply with the required assumptions to apply a statistic parametric method, such as normality and homocedasticity. In general, when there is a violation of one or all the assumptions, researchers usually tend to transform the scale of data. With the use of these transformations, not always the accomplishment of the assumptions of parametric analysis is achieved. It is being proved that its use improves the adjustment of the data, but it can complicate the interpretation of the results due to the change of scale of the variable. Other option is to employ a nonparametric method, which does not require the assumptions but it needs a good experimental design, an adequate sample size and a correct randomization of treatments. Nonparametric methods have a minor power because there are greater chances of do not detect differences among treatments that already exists. The probability to take an incorrect decision is higher if the sample size is small and there are different number of replicas by treatment. There are more complex analyses which require more intensive compute method, for example: methods of resample, randomization, permutation, Bayesian, meta-analysis and generalized linear models. To interpret the results obtained by these methods, the researchers have to change their minds because they are used to work with the methods of frequency statistic.

**Keywords:** Data transformation, Nonparametric methods, Generalized Linear Models, Null Models

**Recibido:** abril 2018 **Aceptado:** junio 2018

Publicado online 25 de agosto de 2018. ISSN 2410-5546 RNPS 2372 (DIGITAL) - ISSN 0253-5696 RNPS 0060 (IMPRESA)

### INTRODUCCIÓN

En el campo de las ciencias experimentales, los investigadores han mostrado, en ocasiones, cierto rechazo al empleo de análisis estadísticos inferenciales en la presentación y discusión de los resultados de sus investigaciones. Este rechazo puede deberse a varios factores, entre los que se pueden destacar, un limitado conocimiento teórico - práctico de los métodos estadísticos básicos o de los más específicos y complejos, pocas habilidades en el manejo de datos o en el trabajo con programas estadísticos, acceso limitado a programas estadísticos reconocidos o a las versiones más actuales

de los mismos, o sencillamente, la creencia de que solo es necesario mostrar los resultados mediante algunos gráficos o tablas, no siempre bien elaborados, en los que básicamente solo aparecen algunos estadísticos descriptivos.

Es real que no todas las investigaciones requieren el desarrollo de experimentos muy complejos o extensos en el tiempo, y en la medida que los mismos sean más o menos sencillos, los métodos de análisis estadísticos que se van a utilizar también lo serán. No obstante, siempre resulta adecuado realizar y mostrar algún tipo de análisis que sea el apropiado al experimento en cuestión y se haya empleado correctamente.

\*Dpto. de Biología Vegetal, Facultad de Biología, Universidad de La Habana, Calle 25, N° 455, Vedado, La Habana, Cuba. C.P. 10 400. e-mail: lene@fbio.uh.cu

Existen algunos factores que pueden atentar contra la calidad del experimento o investigación, y el empleo o no de un correcto análisis estadístico de los resultados. Entre estos factores se pueden destacar: el número de unidades experimentales utilizadas (objeto físico o biológico sobre el cual se aplican los tratamientos y se miden las variables), o el número de réplicas o repeticiones empleadas por cada uno de los tratamientos que se evalúan en el experimento, el tiempo y los recursos destinados a la investigación: reactivos, materiales de laboratorio o de campo, equipamiento, personal capacitado, presupuesto disponible, entre otros. Lamentablemente no siempre se cuenta con la cantidad y calidad necesaria de los mismos para poder desarrollar los experimentos de la mejor forma posible. No obstante, el investigador debe ser capaz de tener en cuenta los posibles sesgos y errores que tiene su investigación, al presentar, discutir y concluir sus resultados.

Muchos de los análisis estadísticos más comúnmente empleados por los investigadores que trabajan en el campo de las ciencias biológicas y agrícolas, como las pruebas t de Student de comparación de medias, el análisis de varianza, las pruebas de comparación múltiple de medias, los análisis de regresión y correlación lineal, entre otros, fueron desarrollados para emplearse en datos que cumplan con una serie de premisas especificadas por el autor de la prueba en cuestión. Esos análisis se agrupan en los métodos estadísticos paramétricos, los cuales tienen cierta preferencia en su utilización con respecto a los métodos no paramétricos, pues tienen una potencia superior, siempre que el tamaño de muestra sea suficientemente grande y se cumplan las premisas establecidas. La mayoría de los análisis paramétricos requieren de los supuestos de normalidad de los datos, independencia de los errores, aleatorización y homogeneidad de varianzas, cuya comprobación es necesaria para sustentar la validez del análisis. El incumplimiento de algunas de estas premisas conlleva a conclusiones no válidas y errores en el proceso de toma de decisiones (Herrera & *al.* 2012).

Con frecuencia se obtienen datos no normales o con varianzas heterogéneas en variables biológicas (Whitlock & Schluter 2009). Por ejemplo: el porcentaje de semillas que germinan se ajusta mejor a una distribución binomial, el conteo de malas hierbas por parcela a una distribución de Poisson o binomial negativa, el tiempo de floración a una distribución exponencial o gamma, la proporción del área de la hoja afectada a una distribución beta, entre otras muchas variables (Stroup 2015).

Los datos provenientes de las investigaciones relacionadas con la genómica, proteómica, transcriptómica, y otras áreas afines, con frecuencia tienen distribuciones no normales (Stroup 2015). En todas las distribuciones,

excepto la normal, la varianza depende de la media, y como consecuencia, cuando no hay normalidad, las varianzas tienden a ser heterogéneas, aumentan con las medias (Zar 2010).

Si se consideran los aspectos previamente mencionados, pueden surgir las siguientes interrogantes: ¿Cómo se analiza el comportamiento de las variables objeto de estudio en el experimento? ¿cómo seleccionar las pruebas estadísticas adecuadas en función de los objetivos de la investigación?, ¿cómo proceder en el caso de que las premisas de las pruebas seleccionadas no se cumplan?, ¿continúa siendo válido el empleo de las transformaciones de escala? A algunas de estas preguntas se les pretende dar respuesta en el presente artículo de revisión.

### **¿Cómo se analiza el comportamiento de las variables objeto de estudio en el experimento?**

Una forma de analizar correctamente los datos obtenidos en una investigación, es identificar la distribución apropiada de los mismos, al realizar un estudio de la variable antes de hacer los análisis (McDonald 2014, Bandera 2018). Para estudiar el comportamiento de las variables se pueden calcular algunos estadísticos descriptivos de tendencia central y dispersión, que son cantidades que capturan características importantes de las distribuciones de frecuencias de los datos (Whitlock & Schluter 2009). Entre las medidas más utilizadas se encuentran, la media aritmética, la moda, la mediana, los valores mínimos y máximos, la varianza, la desviación estándar y el coeficiente de variación. El cálculo de estos valores permite caracterizar la estructura de los datos desde de un punto de vista central y su dispersión con respecto a la media (Zar 2010). También, en ocasiones, posibilita la detección de valores atípicos que se encuentran muy alejados del resto de los valores de la muestra, denominados observaciones atípicas, aberrantes o extremas (*outliers* en inglés) (Whitlock & Schluter 2009, Zar 2010), cuya presencia en la muestra puede afectar los resultados de las pruebas estadísticas.

También se pueden emplear métodos gráficos como los histogramas, los gráficos de caja y bigote, entre otros, para representar la distribución de frecuencias de las variables. Estos gráficos permiten visualizar el comportamiento de las variables, su simetría o asimetría, si se ajustan a la campana de Gauss característica de la distribución normal, la presencia de dos o más modas y de observaciones atípicas. Con los histogramas se puede observar mejor la asimetría de la distribución (Avanza & *al.* 2003) y con los gráficos de caja y bigote se puede visualizar la media, la mediana, los cuantiles, la asimetría y los *outliers* cuando existen (Di Rienzo & *al.* 2005, Garson 2012).

En los programas estadísticos se pueden encontrar varias opciones para calcular estadísticos descriptivos, realizar los gráficos y las pruebas que permiten verificar el ajuste a las premisas de los métodos paramétricos, como las pruebas de Kolmogorov-Smirnov, Shapiro-Wilk, Lilliefors para la normalidad y de Bartlett, Levene para la homocedasticidad (Whitlock & Schluter 2009, Zar 2010). En algunos programas como el SAS (*Statistical Analysis System*) se pueden desarrollar procedimientos como el UNIVARIATE que permite calcular estadísticos descriptivos, detectar las observaciones más extremas, realizar pruebas de Bondad de Ajuste a la normalidad y diferentes tipos de gráficos. Además, existe otro procedimiento denominado SEVERITY con el cual se puede determinar a cuál distribución muestral se ajusta realmente cada una de las variables evaluadas, en caso de no cumplirse la premisa de normalidad (SAS Institute 2012).

### ¿Cómo decidir cuál prueba estadística usar en función de los objetivos de la investigación?

Uno de los mayores retos de los análisis estadísticos, es encontrar entre el gran número de métodos disponibles, el correcto para responder una pregunta particular. En realidad, con la gran disponibilidad de programas estadísticos, el principal reto para los investigadores, es seleccionar el método adecuado para el análisis de sus datos, porque el procedimiento estadístico se realiza con el empleo de un programa en la computadora (Whitlock & Schluter 2009). Pero, ¿cómo se puede seleccionar el método correcto? Afortunadamente, el procedimiento de análisis lógico que se sigue para seleccionar el método correcto es similar en cada caso.

Whitlock & Schluter (2009) formularon varias preguntas que son necesarias responder, para ayudar en la selección del método a utilizar: ¿Se está evaluando una variable o se quiere evaluar la asociación entre dos o más variables?, ¿Las variables son categóricas o numéricas?, ¿Las mediciones están pareadas o son independientes? ¿Cuáles son las premisas de la prueba seleccionada? ¿Los datos cumplirán con esas premisas? A estas preguntas se les puede añadir las siguientes: ¿Qué pruebas se utilizan para verificar el cumplimiento de las premisas? ¿Cómo se debe proceder cuando las premisas no se cumplen? ¿Qué prueba o método alternativo se puede utilizar?

Los diferentes métodos que se pueden aplicar, dependen de si se están evaluando una, dos o más variables simultáneamente porque las pruebas para dos o más variables están dirigidas a analizar la relación entre ellas. Cuando se evalúa la relación entre dos variables, es importante saber si las variables son numéricas, categóricas o una mezcla de ambas. Si se van a comparar dos o más tratamientos, es importante

conocer si las muestras son independientes o están relacionadas o pareadas, porque los tratamientos deben ser aplicados a cada unidad experimental. En estos dos tipos de experimentos, se utilizan diferentes análisis estadísticos (Whitlock & Schluter 2009). También es importante determinar si las mediciones se realizaron de manera repetida en el tiempo sobre las mismas unidades experimentales, lo que se conoce como medidas longitudinales, porque en estas situaciones se incumplen las premisas de los métodos paramétricos y se deben emplear modelos más complejos que ajusten estos datos (Stroup 2015). También se deben conocer las pruebas que se utilizan para verificar el cumplimiento de las premisas de los métodos paramétricos. En caso de su incumplimiento, el investigador debe decidir cómo va a proseguir el análisis y qué pruebas va a utilizar.

### ¿Cómo deben ser analizados los datos que no cumplen con las premisas de los métodos paramétricos?

Existen varias opciones para trabajar con los datos que no cumplen con las premisas. Pocas variables biológicas muestran un ajuste exacto a la normalidad (Whitlock & Schluter 2009). Los datos no normales pueden ser producto de una mezcla de múltiples distribuciones o procesos, o que tengan otro tipo de distribución no normal (McDonald 2014). Primero se debe estar seguro que la no normalidad es debida a la naturaleza de los datos, y no a datos perdidos, no declarados o a errores en la entrada de los mismos que se pueden corregir.

Otras razones de no normalidad son la propia naturaleza de la variable y la presencia de valores que son muy extremos en relación con el resto de los datos (*outliers*) (Osborne 2002). Existe un gran debate en la literatura con relación a si estos valores atípicos deben ser eliminados o no por la influencia que tienen en los resultados de los análisis. Autores como Osborne (2002) y Judd & *al.* (2009) están a favor de su eliminación pues aseguran que es honesta, deseable e importante. En ocasiones, una observación atípica puede haberse obtenido en una unidad experimental que no pertenece a la población en estudio, es un individuo de otra especie o tiene una edad muy diferente de la del resto de las unidades de la muestra; o puede deberse a errores que se cometieron durante la conducción del experimento. Cuando el investigador está seguro que la observación atípica tuvo alguna de estas causas, las puede eliminar (Zar 2010). Sin embargo, otros autores como Orr & *al.* (1991), no están de acuerdo con la eliminación de estos valores. Algunos *outliers* pueden ser fenómenos interesantes que pueden conllevar al descubrimiento de un conocimiento inesperado u observaciones correctas obtenidas por azar en la población, por lo que su eliminación puede conducir a la pérdida de información útil y la obtención de resultados inexactos o incorrectos en los análisis estadísticos (Cheng 2000, Zar 2010). No es apropiado

descartar un valor porque parece razonablemente extremo. Si el investigador está claro de que la causa del *outlier* es un error puede corregirlo o eliminarlo (Whitlock & Schluter 2009, Zar 2010). Murphy & Lau (2008) sugirieron que, si en la investigación no se encuentra una causa probable del valor atípico, se puede realizar un análisis de los datos con el valor atípico y sin él, y si las conclusiones son diferentes, entonces se considera que el valor atípico tiene influencia y se debe mostrar en los resultados.

En las últimas décadas se han desarrollado varios métodos para la detección y el manejo de los *outliers*. No obstante, en ocasiones, aunque se remueva el *outlier* se mantiene la no normalidad, y la transformación de los datos es una opción viable para mejorar el ajuste (Osborne 2002).

Antes de 1990, la pregunta: ¿cómo se deben analizar los datos no normales? parecía estar resuelta. El Teorema del Límite Central garantizaba que, a pesar de la distribución de los datos en la población, si se tiene un número suficientemente grande de observaciones, la distribución muestral de las medias se puede asumir como aproximadamente normal. Esto se debe a la robustez del análisis de varianza (ANOVA), resumida en un excelente artículo por Miller (1997). En los casos de muestras más pequeñas con falta de normalidad y de homocedasticidad, se ha tendido siempre a realizar transformaciones de la escala de los datos que permiten estabilizar la varianza, o a emplear métodos no paramétricos (Stroup 2015).

Entre 1990 y 2010, los avances en la teoría y la metodología estadística que se habían ido incubando por décadas, se unen con el rápido y sostenido incremento de las capacidades de cálculo, y con el desarrollo de computadoras que pudieran ser programadas con instrucciones escritas, lo que hizo posible la creación de programas estadísticos modernos como el SAS (SAS Institute 2012). La teoría de los modelos lineales generales, en la cual estaban basados los primeros programas estadísticos, no satisfacía los modelos de efectos mixtos, experimentos en parcelas divididas y datos no normales. El avance fundamental de este período se refiere principalmente al empleo de los Modelos Lineales Generalizados Mixtos (GLMM, por sus siglas en inglés, *Generalized Linear Mixed Models*) (Stroup 2015).

Los GLMM permiten que la media de una población dependa de un predictor lineal a través de una función de enlace (*link function*) de tipo no lineal, que varía en dependencia de la distribución probabilística a la cual se ajuste la variable respuesta, la cual debe pertenecer a la familia exponencial (Wang & al. 2015), que comprende

a la mayoría de las distribuciones discretas y continuas. Aunque estos modelos fueron diseñados inicialmente solo para miembros de la familia exponencial, se han extendido a un rango mucho más amplio de aplicaciones, al utilizar métodos de estimación de cuasi-verosimilitud (Littell & al. 2006).

La estimación de parámetros en los modelos lineales generalizados se realiza mediante los procedimientos de máxima verosimilitud (ML, *Maximum Likelihood*) y máxima verosimilitud restringida (REML, *Restricted Maximum Likelihood*) (SAS Institute 2012). La verosimilitud mide cuan bien los datos producen un valor particular para un parámetro, es la probabilidad de obtener un valor particular para un parámetro si este se iguala a ese valor. Cuando un parámetro se estima por máxima verosimilitud el valor que se obtiene es el que producen con mayor probabilidad o verosimilitud los datos, o sea es el mejor estimado del parámetro (Oehlert 2012). Los estimadores de máxima verosimilitud restringida maximizan sólo la porción de la verosimilitud que no depende de los efectos fijos, por lo que son una versión restringida de los estimadores ML, y se prefieren como método para analizar grupos de datos grandes, con estructuras complejas y diseños desbalanceados (De Farias Neto & Resende 2001).

Los modelos lineales generalizados ofrecen nuevas posibilidades porque hacen posible extender los modelos lineales clásicos de efectos fijos, al incluir efectos aleatorios y los Mejores Predictores Lineales Insesgados (BLUP) que permiten estimar los efectos aleatorios en un modelo mixto por el método de máxima verosimilitud restringida (Piepho & al. 2008). Permiten procesar datos no normales, con varianzas heterogéneas, que estén correlacionados y estimar sus componentes de varianza asociados al modelo. Los procedimientos de estimación que utilizan posibilitan reducir los sesgos cuando los datos están incompletos, desbalanceados o ajustar datos dispersos. Además, permiten modelar la estructura de los errores en datos provenientes de mediciones longitudinales (Stroup 2012).

A mediados de los años 2000, comenzaron a aparecer programas comerciales y libres que permitían desarrollar los GLMM, como los paquetes GLMPQL, GEE y LME4 del programa R, y los procedimientos GLIMMIX (*Generalized Linear Mixed Models*) y GENMOD (*Generalized Models*) del paquete SAS (Stroup 2015). La diferencia entre los dos procedimientos del SAS radica en que con el procedimiento GENMOD solo se puede procesar modelos generalizados de efectos fijos, no modelos mixtos como con el procedimiento GLIMMIX.

Este fue un momento crítico para el análisis estadístico. Por primera vez, existían paquetes de programas

útiles para implementar un gran rango de modelos estadísticos que trataban de acomodar experimentos complejos y datos no normales. El extensivo desarrollo de la teoría y la metodología acumulado durante las décadas previas, estuvo disponible entonces para los investigadores, de una forma accesible. Esto explica por qué la interrogante de: ¿cómo hacer análisis estadísticos con datos no normales?, está siendo ahora valorada nuevamente. La estadística, al igual que otras disciplinas es dinámica, no es un grupo fijado de reglas incambiables (Stroup 2015).

El empleo de los GLMM requiere de un cambio en la mentalidad de los investigadores. Los hábitos adquiridos al trabajar con el ANOVA, implican transformar los datos no normales, lo cual no necesariamente ayuda y con frecuencia impide trabajar de forma efectiva con los modelos generalizados. Vale la pena preguntarse entonces: ¿Se debe ignorar el incumplimiento de las premisas de los métodos paramétricos? ¿Todavía son relevantes las transformaciones de los datos? ¿Cuándo se deben considerar esenciales los modelos generalizados mixtos?

Whitlock & Schluter (2009) sugirieron tres posibles alternativas para analizar variables biológicas que no cumplen con las premisas: (1) Ignorar el no cumplimiento de las premisas, (2) Transformar los datos y (3) Emplear un método no paramétrico. Las tres alternativas asumen que los datos muestrales son aleatorios e independientes. Por su parte, Avanza & *al.* (2003) plantearon que ante situaciones donde existen fuertes dudas sobre el incumplimiento de las premisas se pueden seguir varios caminos: (1) si las desviaciones en las hipótesis de cumplimiento no son muy grandes, se puede aplicar la técnica paramétrica teniendo en cuenta las faltas en que se incurre y que la técnica es aproximada; (2) se pueden utilizar técnicas alternativas que no exijan el cumplimiento de premisas como métodos no paramétricos, Mínimos Cuadrados Ponderados o Modelos Lineales Generalizados; (3) aplicar transformaciones a los datos que logren solucionar las desviaciones de las premisas; (4) en algunas situaciones se puede simplificar el modelo, al reducir el orden o eliminar interacciones. Se pueden utilizar también otras alternativas, por ejemplo: las simulaciones, los métodos de remuestreo, permutación y aleatorización que necesitan de un gran número de cálculos (Whitlock & Schluter 2009).

#### **¿Cuándo se puede ignorar el incumplimiento de las premisas de un análisis?**

En algunas situaciones, se realizan análisis paramétricos cuando los datos no cumplen las premisas. Los métodos que estiman y evalúan medias, no son muy sensibles a las premisas de normalidad y homocedasticidad

(Whitlock & Schluter 2009). Bajo ciertas condiciones algunos métodos como el ANOVA son robustos, o sea, trabajan bien cuando no se cumplen las premisas, si el tamaño de la muestra es grande y el incumplimiento de la premisa no es muy drástico (Zar 2010), o sea, si el valor de probabilidad asociado al estadístico de la prueba no es mucho menor que el nivel de significación ( $\alpha$ ) establecido. Se dice que un procedimiento estadístico es robusto cuando los resultados que brinda no son sensibles al incumplimiento de las premisas del método (Whitlock & Schluter 2009). Sin embargo, una desviación significativa de las premisas puede incrementar seriamente las posibilidades del investigador de cometer error tipo I o tipo II, en dependencia de la naturaleza del análisis, lo que implica resultados inexactos e interpretaciones incorrectas en las pruebas estadísticas (Zou & Kimbeng 2010).

Los intervalos de confianza para la media, la prueba t de comparación de medias y el ANOVA pueden ser usados, en ocasiones, en datos con distribuciones no normales (Zar 2010). Esto es debido a que son pruebas robustas, lo cual se deriva del Teorema del Límite Central que plantea que una variable que no se distribuye normalmente en la población, puede tener una distribución normal en la muestra si esta es suficientemente grande (Stroup 2015). Los análisis de factores, la correlación lineal, la regresión de mínimos cuadrados y otras técnicas lineales relacionadas, son relativamente robustas ante desviaciones no extremas de la normalidad, pues proveen errores que no son severamente asimétricos (Garson 2012). Sin embargo, los intervalos de confianza para la varianza y la desviación estándar son muy sensibles a la falta de normalidad y los métodos que los determinan no son robustos al incumplimiento de esta premisa (Whitlock & Schluter 2009). Estos autores también plantearon que los métodos de comparación de medias son los robustos, no los métodos similares a la prueba F de comparación de varianzas, la cual no es robusta para el incumplimiento de la premisa de normalidad.

La prueba t de Student y los intervalos de confianza para la media basados en la distribución t, pueden ser utilizados en datos que no cumplen con la premisa de normalidad, cuando las muestras son grandes, incluso si las distribuciones de los datos en las dos muestras son asimétricas (Fagerland 2012). Zar (2010) planteó que los efectos de la no normalidad sobre el empleo de la prueba t, son menores si las muestras tienen igual tamaño y las varianzas son similares. Fagerland (2012) no recomienda el empleo de métodos no paramétricos, como la prueba de Wilcoxon-Mann-Whitney, en muestras grandes porque la prueba es sensible a la heterogeneidad de varianzas, y puede detectar diferencias entre las medias o las medianas

con una probabilidad menor que el 5%, basada en las diferencias en dispersión de las dos distribuciones aun cuando las medias o medianas no difieren.

Sin embargo, al investigador puede surgirle otra interrogante, ¿cuán grande debe ser una muestra para ignorar la premisa de normalidad? Whitlock & Schluter (2009) sugieren que la respuesta depende de la forma de la distribución porque si se están comparando dos grupos y ambos difieren de la normalidad en diferentes formas, entonces, aun sutiles desviaciones de la normalidad pueden causar errores en los análisis, incluso con muestras grandes. Estos autores plantearon que, si se comparan dos grupos y ambos tienen la misma distribución asimétrica a la derecha, se pueden obtener resultados confiables con la prueba t de Student, con 30 unidades muestrales en cada grupo.

Sin embargo, si se comparan dos grupos con distribuciones diferentes y diferente asimetría, se requieren muestras de 500 o más unidades muestrales, para obtener resultados válidos con una prueba t. Si las distribuciones son más diferentes que el caso anterior, entonces se debe evitar realizar la prueba t de comparación de medias, y se debe emplear un método alternativo. Si la distribución de frecuencias es muy asimétrica, la prueba t no va a brindar resultados confiables, incluso con tamaños de muestra extremadamente grandes. Las distribuciones de frecuencias que contienen *outliers*, no se deben analizar con una prueba t o con intervalos de confianza basados en la distribución t porque los métodos que necesitan del ajuste a la normalidad son muy sensibles a la presencia de *outliers* (Whitlock & Schluter 2009). Sin embargo, Fagerland & Sandvik (2009) sugieren que cuando el tamaño de la muestra es de 200, la prueba t es robusta aún con distribuciones muy asimétricas.

Otra premisa importante en los métodos paramétricos es la homogeneidad de varianzas en las muestras a comparar. Se puede ignorar la presencia de heterocedasticidad (varianzas heterogéneas), cuando se van a comparar medias muestrales entre dos grupos con el empleo de la prueba t de Student, si los tamaños de las dos muestras son mayores de 30 unidades muestrales, y aproximadamente iguales. Si los tamaños muestrales no son iguales, o si la diferencia entre las desviaciones estándar es muy grande, se debe emplear la prueba t' o aproximación de Welch (Fagerland 2012). Si además no se cumple la premisa de normalidad, entonces se deben transformar los datos, emplear un método no paramétrico u otro método alternativo. El ANOVA es más sensible a la falta de homocedasticidad que de normalidad, pues la probabilidad de cometer error tipo I se incrementa mientras más heterogéneas sean las varianzas y diferentes los tamaños de muestra (Zar 2010).

Como no existen instrucciones definitivas de cómo actuar en cada caso, se debe hacer un estudio de la variable antes de realizar cualquier prueba, como se mencionó anteriormente. Cuando las distribuciones están muy distantes de la normalidad, tienen *outliers*, o las distribuciones de los grupos a comparar son muy diferentes y tienen varianzas muy heterogéneas, no se debe ignorar el incumplimiento de las premisas. En estos casos se debe utilizar una vía alternativa como transformar la escala de la variable, emplear un método no paramétrico, de simulación o un modelo lineal generalizado.

### **Empleo de transformaciones de escala**

Las transformaciones de escala constituyen un aspecto importante en la preparación de los datos para su análisis estadístico. Son herramientas comúnmente utilizadas en el análisis de datos cuantitativos para mejorar el ajuste de las variables a la normalidad, lograr la homogeneidad de varianzas o la independencia de errores (Osborne 2010). Sin embargo, Zar (2010) planteó que no compensan la ausencia de aleatorización.

Las transformaciones son funciones matemáticas que se aplican a los valores de la variable (Osborne 2010). Cuando se transforman los datos, se realiza la misma operación matemática para cada observación, y se utilizan los valores transformados para ejecutar la prueba estadística en cuestión (McDonald 2014); por lo que se comparan las medias transformadas. Las transformaciones cambian la naturaleza fundamental de los datos y la interpretación de los resultados (Osborne 2002). Después de realizar el análisis estadístico con los datos transformados, no es recomendable presentar las medias, desviaciones estándar, errores estándar y otros estadísticos descriptivos, en las unidades transformadas, debido a que el cambio de escala de la variable puede resultar poco informativo al mostrar los resultados en un gráfico (Osborne 2010, Manikandan 2010). Se deben retransformar los resultados con el empleo de la función matemática inversa a la empleada en la transformación, para presentar los resultados (McDonald 2014). Al transformar los datos, todos los cálculos se deben hacer en la misma escala y la retransformación solo puede ser realizada al final. Se debe mencionar la transformación empleada y la razón por la cual se seleccionó (Manikandan 2010).

Existe un gran número de transformaciones a utilizar. Se debe escoger una que sea comúnmente empleada en el campo de estudio según el tipo de variable evaluada (McDonald 2014). Es importante decidir qué transformación utilizar antes de realizar el análisis estadístico. No se debe probar con cada transformación, buscando cuál da un resultado significativo de la prueba a realizar, porque las transformaciones se emplean

para optimizar los datos, no para optimizar el resultado del análisis estadístico. Si se tiene un gran número de observaciones, se pueden comparar los efectos de diferentes transformaciones sobre la normalidad y la homocedasticidad de la variable. Si se tiene un pequeño número de observaciones, no se va a observar un gran efecto de las transformaciones sobre la normalidad y la homocedasticidad (McDonald 2014).

Existen una gran variedad de posibles transformaciones, desde adicionar una constante hasta multiplicarla, elevar los valores al cuadrado, calcular la raíz cuadrada, el logaritmo, el inverso o emplear una transformación trigonométrica como el arco seno ( $\arcsen$ ) (Osborne 2002, 2010). Algunos autores como Osborne (2002, 2008) y McDonald (2014), han planteado algunas recomendaciones que se pueden tener en cuenta al emplear las transformaciones tradicionales, como anclar el valor mínimo de la distribución en uno porque la eficiencia de algunas transformaciones se ve afectada cuando los valores se dispersan por debajo de este valor.

*La transformación logarítmica ( $\log Y$  o  $\ln Y$ )* es una de las más empleadas en las ciencias biológicas, pues muchas variables biológicas tienen distribuciones log-normales, lo que significa que después de aplicar la transformación logarítmica, los valores se distribuyen normalmente (Feng & al. 2014). Esto es debido a que, si se toma un grupo de factores independientes y se multiplican todos juntos, el producto resultante es log-normal (Manikadan 2010, McDonald 2014). Se utiliza principalmente cuando existe proporcionalidad entre la media y la desviación estándar, la distribución de frecuencias es asimétrica a la derecha y existen efectos multiplicativos en el modelo lineal (Whitlock & Schluter 2009, Zar 2010). Se debe aplicar en valores mayores que cero, si en los datos se incluye el cero, se debe emplear  $\log(Y + 1)$ . Para retransformar los datos se utiliza el antilogaritmo. No siempre la transformación logarítmica resuelve el problema, puede que modifique la forma de la distribución hacia una más simétrica o la haga más asimétrica que con los datos originales (Feng & al. 2014).

*La transformación angular ( $\arcsen \sqrt{p}$ )*, se emplea en datos que se aproximan más a una distribución binomial, como los conteos o porcentajes, los cuales se deben convertir en proporciones antes de aplicarles la transformación (Zar 2010). Las proporciones tienden a no estar normalmente distribuidas, principalmente si la media está cercana a cero o a uno, y las desviaciones estándar de los grupos son heterogéneas. Con frecuencia se logra el cumplimiento de las dos premisas con su empleo. La transformación inversa es elevar al cuadrado el seno del valor transformado (Whitlock & Schluter 2009).

La transformación raíz cuadrada ( $\sqrt{Y}$ ) es útil cuando los datos son conteos, que se aproximan más a la distribución de Poisson, y cuando existe proporcionalidad entre la media y la varianza de los grupos a comparar (Zar 2010). Cuando los valores son muy pequeños o se incluyen ceros, se debe agregar una constante a cada uno de los valores, que puede ser: 0,5 o 3/8, antes de calcular la raíz (Zar 2010). Si existen valores negativos, se debe adicionar un valor constante que los eleve por encima de uno (Osborne 2002, 2010). Para retransformar se elevan al cuadrado los valores (McDonald 2014).

*La transformación inversa o recíproca ( $1/Y$ )* se utiliza cuando las desviaciones estándar de los grupos son proporcionales a los cuadrados de las medias y las distribuciones tienen una asimetría positiva (Zar 2010). Si existen valores cero, se calcula:  $1/(Y+1)$ . Esta transformación tiene el efecto de revertir el orden de los valores, o sea, convertir los mayores en los menores y viceversa. Se pueden reflejar o revertir los valores antes de realizar la transformación. Para ello, se multiplican por -1 los valores, se adiciona una constante para anclar el valor mínimo en uno, y entonces se aplica la transformación, en orden de que los valores sean idénticos a los datos originales (Osborne 2010).

*La transformación cuadrada ( $Y^2$ )* se puede utilizar cuando la distribución de frecuencias de los datos es asimétrica a la izquierda y las desviaciones estándar disminuyen en la medida que se incrementan las medias (Zar 2010). Si esta no funciona, se puede probar con la *transformación antilogaritmo ( $e^Y$ )* (Whitlock & Schluter 2009).

La mayoría de las transformaciones previamente mencionadas pertenecen a una clase denominada transformaciones de potencia. *Las transformaciones de Box y Cox* son una familia de transformaciones de potencia continuas que utilizan el método de máxima verosimilitud, para determinar la transformación de potencia óptima (Garson 2012) con el fin de lograr corregir sesgos en la distribución de errores, varianzas desiguales y principalmente para corregir la no linealidad en la relación. La transformación mejora el ajuste a la normalidad y la homocedasticidad en distribuciones asimétricas a la derecha y a la izquierda (Osborne 2010). Esta familia de transformaciones utiliza el parámetro  $\lambda$  y está definida solo para variables con valores positivos (Avanza & al. 2003), por lo que, si se tienen valores negativos, se debe adicionar una constante para hacerlos positivos. El parámetro  $\lambda$  puede tomar un número infinito de valores, por lo que teóricamente, se puede calibrar la transformación para que sea máximamente efectiva en lograr el ajuste a la normalidad (Osborne 2010).

Un beneficio del uso de las transformaciones es que cuando se logra el cumplimiento de una de las premisas, es frecuente acercarse al cumplimiento de otra (Manikadan 2010). Por ejemplo, el empleo de la transformación raíz cuadrada, puede ayudar a igualar un grupo de varianzas, y como esto comprime el límite superior de la distribución más de lo que comprime el límite inferior, puede tener un efecto en hacer más cercanas a la forma normal, las distribuciones con asimetría positiva. Si se decide transformar, es importante comprobar si la variable alcanzó la normalidad o está cercana a esta después de la transformación, o sea, comprobar si funciona.

Algunas transformaciones trabajan para comprimir el lado derecho de la distribución más que el izquierdo, por lo que son efectivas cuando hay una asimetría positiva. Cuando se tiene una distribución con asimetría negativa, se debe adicionar una constante que acerque los valores a uno, y entonces aplicar la transformación, para después restaurar el orden original de la variable (Osborne 2002). Este autor planteó que todas las transformaciones reducen la no normalidad al reducir el espacio relativo entre los valores en el lado derecho de la distribución más que los valores del lado izquierdo, o sea tienen un efecto de comprimir los números mayores mucho más que los números menores; pero esto no sucede así cuando la distribución se aleja de uno.

Algunos investigadores encuentran difícil la idea de aceptar la transformación de los datos. Es importante reconocer que las conclusiones que se obtienen de los datos transformados no siempre transfieren cuidadosamente las mediciones originales. Grissom (2000) planteó que las medias de las variables transformadas pueden ocasionalmente invertir la diferencia entre las medias de las variables originales. Si se está dispuesto a aceptar que es permisible transformar un grupo de mediciones en otro, entonces varias posibilidades se hacen disponibles para modificar los datos, para ajustarlos más a las premisas de las pruebas estadísticas paramétricas.

Howell (2007) sugirió observar las medias transformadas y sin transformar, y asegurarse de que ambas muestran la misma información. Además, planteó que no se deben convertir las desviaciones estándar. Tabachnick & Fidell (2007) plantearon que, aunque las transformaciones de los datos son recomendadas como un remedio para los *outliers* y la falta de normalidad, linealidad y homocedasticidad, no son recomendadas universalmente, debido a que un análisis es interpretado por las variables que contiene, y las variables transformadas son a veces difíciles de interpretar.

Aunque las transformaciones constituyen una opción importante para el analista, modifican la naturaleza de la variable, lo que hace más difícil su interpretación (Osborne 2010). Los investigadores deben ser cuidadosos cuando interpretan datos transformados porque no siempre todos los datos permanecen en el mismo orden relativo que antes de la transformación, ni la distancia entre los valores adyacentes continúa siendo la misma porque la transformación altera la igualdad de distancia, y la alteración es diferente según la transformación.

La popularidad de las transformaciones de escala ha conducido a un mal uso de las mismas y a interpretaciones incorrectas de los resultados experimentales, lo cual es un problema común en varios métodos estadísticos básicos (Feng & al. 2014). Se deben utilizar las transformaciones de datos con cuidado y nunca a menos que la razón este muy clara. La transformación puede alterar la naturaleza de los datos, como cambiar su escala de medición de intervalo o razón a ordinal, y crear relaciones curvilíneas, de complicada interpretación (Osborne 2002). Además, puede hacer más asimétrica la distribución de los datos (Feng & al. 2014). Es posible que el cambio de escala ayude, pero todavía exista la presencia de *outliers*. El próximo paso es analizar todos los datos con y sin *outliers*. Si los resultados son los mismos con estos dos análisis paralelos, se pueden eliminar los *outliers* o mantenerlos cuando se está seguro de que pertenecen a la población muestreada, pero si se obtienen resultados diferentes, se deben obtener más datos, lo cual no es siempre posible.

Para varias aplicaciones, más que tratar de encontrar una distribución estadística apropiada o transformar a un modelo los datos observados, es probable que sea mejor abandonar este enfoque clásico y cambiarlo a métodos modernos libres de distribución. Por ejemplo: un enfoque que puede evitar muchos de esos problemas es el de las ecuaciones de estimación generalizadas (*GEE* de sus siglas en inglés, *Generalized Estimating Equations*). Este enfoque ignora la premisa de la distribución y provee una inferencia válida a pesar de la distribución de los datos (Feng & al. 2014). Sin embargo, este es solo apropiado para datos asimétricos, si los datos son razonablemente modelados por una distribución paramétrica como la distribución normal, es preferible utilizar los métodos estadísticos clásicos porque estos generalmente proveen una inferencia más eficiente que las *GEE* (Stroup 2015).

### **Empleo de métodos no paramétricos**

Si no se pueden desarrollar los análisis paramétricos al ignorar la desviación de las premisas o al transformar los datos, se puede recurrir a un método no paramétrico homólogo al paramétrico de interés. Estos métodos asumen menos premisas acerca de la distribución probabilística de la variable de interés, en la población



de la cual se tomaron los datos, por lo que se pueden emplear casi siempre, pero son menos potentes que los métodos paramétricos. Son muy útiles cuando se tienen *outliers* en la distribución, porque generalmente trabajan con rangos, los cuales no son afectados por estos (Whitlock & Schluter 2009, Zar 2010).

Cuando se comprueba que se cumplen las premisas de una prueba específica, la probabilidad de cometer error tipo I es igual al valor de  $\alpha$ , tanto para los métodos paramétricos como los no paramétricos. Sin embargo, cuando las premisas no se cumplen, entonces la probabilidad de error tipo I se hace mayor que el valor de  $\alpha$  especificado. Este incremento de la tasa de error tipo I es la razón por la cual no se deben usar los métodos paramétricos cuando las premisas no se cumplen. Bajo estas condiciones, los métodos no paramétricos conducirán a una probabilidad de error tipo I igual a  $\alpha$ . Sin embargo, con la probabilidad de error tipo II (fallar en rechazar una hipótesis nula incorrecta) no sucede lo mismo. Como los métodos no paramétricos solo usan los rangos, utilizan menos información de los datos que los métodos paramétricos. Esta pérdida de información causa que los métodos no paramétricos tengan menos potencia que el correspondiente método paramétrico. Cuando las premisas se cumplen, tienen una menor probabilidad de rechazar una hipótesis nula falsa y una mayor tasa de error tipo II, que el método paramétrico (Whitlock & Schluter 2009).

Por esta razón, algunos investigadores prefieren emplear los métodos no paramétricos solo cuando las premisas no se cumplen, ni después de transformar la escala de los datos. No obstante, la menor potencia de los métodos no paramétricos no es de gran importancia cuando las premisas son fuertemente incumplidas, pues en esos casos no se debe emplear la prueba paramétrica (Whitlock & Schluter 2009).

Los métodos no paramétricos también tienen sus premisas. Aunque no tienen en cuenta la distribución probabilística de los datos, si asumen que las muestras a comparar fueron seleccionadas aleatoriamente de sus respectivas poblaciones y que los errores son independientes y aditivos. Si el muestreo no se realizó de forma aleatoria, el método no paramétrico dará resultados erróneos de manera similar al método paramétrico (Whitlock & Schluter 2009).

La prueba de Wilcoxon de los rangos con signos asume que la distribución de mediciones en la población es simétrica, lo que limita la utilidad del método. Por otra parte, la prueba U de Mann-Whitney es sensible a la no similitud en la forma de la distribución de los grupos que compara, lo que implica que no tengan la misma varianza y asimetría, y puede detectar diferencias entre

medias que no existen y que se deben a las diferencias en las desviaciones estándar de los grupos (Fagerland 2012). La probabilidad de error tipo I (rechazar una hipótesis nula correcta) se hace muy diferente de  $\alpha$ , si las dos distribuciones que se comparan tienen varianzas o asimetrías diferentes (Zar 2010). Esta limitación de la prueba U es poco conocida, por lo que se puede encontrar un uso erróneo de la misma en la literatura científica (Whitlock & Schluter 2009).

Existe un método no paramétrico de Brunner-Munzel, alternativo a la prueba de Wilcoxon-Mann-Whitney de comparación de dos grupos o tratamientos, que ajusta las varianzas heterogéneas, pero no está ampliamente disponible en los programas estadísticos (Fagerland & Sandvik 2009). La utilidad de la prueba de Wilcoxon-Mann-Whitney como alternativa a la prueba t de Student, cuando los datos no se ajustan a la normalidad, es más pronunciada si los tamaños de muestra son pequeños, no se recomienda su uso en muestras grandes (Fagerland 2012).

Los análisis de varianza no paramétricos se utilizan especialmente cuando los k grupos que se comparan no provienen de poblaciones con distribución normal de la variable respuesta (Krutchkoff 1998). También se pueden utilizar cuando las distribuciones de las muestras tienen dispersiones y formas poco diferentes, pero si las varianzas son diferentes la probabilidad de cometer error tipo I se separa más del nivel de significación de la prueba, en concordancia con la magnitud de las diferencias (Zimmerman 2000). La prueba de Kruskal-Wallis (ANOVA de clasificación simple por rangos) tiende a ser más potente si los tamaños de las muestras son más grandes y la potencia disminuye si los tamaños de muestra no son iguales. Cuando las varianzas son heterogéneas, la tasa de error tipo I se afecta menos con la prueba de Kruskal-Wallis que, con el análisis de varianzas paramétrico, si los grupos con las mayores varianzas tienen tamaños pequeños (Zar 2010).

#### **Empleo de otros métodos más complejos. Modelos nulos.**

La estadística de modelos nulos es aquella que utiliza simulaciones computarizadas para generar, a partir de la propia muestra de datos, una distribución de probabilidades en la cual no se tenga en cuenta el posible efecto que se quiere detectar, y que es empleada para ver cuánto se alejan los efectos detectados de aquellos que aparecerían en una distribución aleatoria sin dicho efecto. Gotelli & Graves (1996) definieron un modelo nulo como un modelo generador de patrones que se basa en la aleatorización de datos ecológicos o un muestreo aleatorio de una distribución conocida o imaginaria. Ciertos elementos de los datos se mantienen constantes y otros se permiten variar estocásticamente para crear

nuevos patrones de asociación. La aleatorización es diseñada para producir un patrón que se debía esperar en la ausencia de un mecanismo particular. Dicho en palabras simples, un modelo nulo es, dado un mecanismo de interés, el modelo de cómo debían ser los datos en ausencia dicho mecanismo (Gotelli 2001). Los métodos de modelos nulos se derivan de tres grupos de algoritmos fundamentales: las simulaciones de Monte Carlo, las pruebas de permutación y de aleatorización. Son métodos de cálculo intensivo, particularmente útiles cuando no se cumplen las premisas de las pruebas estadísticas clásicas o cuando no existe un método estándar para realizar el análisis. Estos métodos requieren de un gran número de cálculos por lo que no se pueden realizar sin la ayuda de la computadora. El valor de estas técnicas es que se pueden aplicar en casi todos los tipos de problemas estadísticos (Whitlock & Schluter 2009).

Los métodos de simulación se utilizan en las pruebas de hipótesis, donde su mayor reto es determinar la distribución de probabilidad de una prueba estadística cuando la hipótesis nula es verdadera. El método emplea la computadora para simular o imitar el proceso de muestreo repetido de una población, para aproximar la distribución nula de la prueba estadística. Con la simulación, se crea una población imaginaria cuyos parámetros tienen como valores aquellos especificados por la hipótesis nula. Cuando se simula el muestreo de la población imaginaria con la computadora, se utiliza siempre el mismo protocolo, de la misma forma que se hace cuando se colectan datos reales. El proceso de simulación se repite muchas veces y siempre se calcula el estadístico de la prueba. La distribución de frecuencias de los valores del estadístico de la prueba, que se obtiene de todas las simulaciones, es aproximadamente la distribución nula de la prueba de hipótesis (Whitlock & Schluter 2009).

Los métodos de Monte Carlo son técnicas utilizadas para evaluar como un estadístico puede ejecutarse bajo un muestreo repetido. En estos métodos, la computadora utiliza un número aleatorio de simulaciones que imitan a la población estadística. Los métodos más conocidos son el Bootstrap y el Jackknife, que se utilizan ampliamente para obtener estimados de parámetros de distribuciones probabilísticas desconocidas (Efron & Tibshirani 1993, Mainly 2007).

El método de Jackknife, originalmente introducido por Quenouille en 1949 como método de reducción de sesgos, fue generalizado por Tukey en 1958 como herramienta general para estimar valores de estadísticos y sus intervalos de confianza. La técnica se basa en recalcular ( $n - 1$ ) veces el estadístico en cuestión analizado, dejando fuera cada vez, uno de

los datos originales. Los pseudovalores obtenidos se pueden promediar y obtener intervalos de confianza o cualquier estimado de variabilidad. Es un procedimiento paramétrico que permite reducir el sesgo en el estimado del valor poblacional para un estadístico, y proporciona su error estándar (Efron & Tibshirani 1993).

La técnica Bootstrap fue descrita por Bradley Efron, como una alternativa al Jackknife. Se utiliza para estimar parámetros, probar hipótesis, obtener errores estándar, determinar intervalos de confianza asimétricos, entre otros. Tiene un enorme potencial, sobre todo para biólogos con habilidades de programación. El procedimiento emplea la computadora para crear la distribución muestral, a partir de la selección repetida de muestras aleatorias de los propios datos, por lo cual es un método de remuestreo. Se remuestran los datos para crear varios nuevos grupos de datos, a partir de los cuales se infiere la distribución muestral del estimado. Resulta muy útil este método cuando no se tiene disponible una fórmula para el error estándar o cuando la distribución muestral del estimado no se conoce (Whitlock & Schluter 2009).

En las pruebas de permutación, en lugar de comparar el valor actual del estadístico probado con su distribución estadística estándar, la distribución de referencia es generada por los propios datos. La permutación provee un acercamiento eficiente a las pruebas de hipótesis, cuando los datos no cumplen las premisas de normalidad y homocedasticidad de la prueba estadística que se quiere emplear. Sin embargo, no resuelve los problemas de independencia de las observaciones. Estas pruebas generan pseudomuestras nulas a partir de la determinación de todas las posibles combinaciones de muestras, de igual tamaño a las originales, en un conjunto de datos o matrices, para lo cual se siguen procedimientos de permutaciones. Una permutación es cada una de las ordenaciones posibles de los elementos de un conjunto finito (Anderson & Legendre 1999).

Las pruebas de aleatorización se basan en la reconstrucción de nuevas muestras a partir de la aleatorización de los datos existentes. Constituyen un enfoque muy usado en el estudio de las relaciones entre las especies y variables ambientales (Mainly 2007). Son utilizadas para probar hipótesis de que dos variables están asociadas, como las tablas de contingencia y el análisis de correlación simple, o permiten comparar dos grupos como las pruebas t de Student y Wilcoxon-Mann-Whitney, cuando no se cumplen los requerimientos de estos métodos o cuando no se conoce la distribución nula. Son más potentes que las pruebas no paramétricas basadas en rangos (Whitlock & Schluter 2009). Para poder aplicar estas pruebas se debe trabajar con muestras aleatorias

tomadas de las poblaciones y las distribuciones de los datos deben tener la misma forma en cada población para poder comparar medias o medianas.

### CONSIDERACIONES FINALES

Antes de realizar alguna prueba estadística se deben examinar los datos, en busca de posibles errores cometidos en la entrada de los mismos a la base de datos, valores faltantes, posibles *outliers*, y presencia de asimetría.

Es importante conocer las premisas de la prueba a utilizar, incluso si es un método no paramétrico, los cuales requieren también el examen de los datos y de haber aplicado un correcto diseño experimental.

Aunque se pueden realizar pruebas paramétricas, en datos que no cumplen las premisas de tales análisis y no se transforman, cuando los tamaños muestrales son suficientemente grandes, resulta difícil establecer cual tamaño de muestra se considera suficientemente grande. Las transformaciones de escala, aunque han sido una herramienta universal que se mantiene válida en la actualidad, no siempre se logra con su aplicación el cumplimiento de una o todas las premisas del análisis, cuyo incumplimiento en ocasiones, también sesga los resultados al analizar los datos con métodos no paramétricos.

En experimentos más complejos que requieren el uso de diseños experimentales desbalanceados y con medidas repetidas, no se sugiere el uso de transformaciones de escala para datos no normales. Se recomienda el empleo de modelos lineales generalizados.

En la actualidad también se están implementando otros métodos alternativos como los meta-análisis, los modelos nullos, la modelación y la estimación de parámetros con el empleo de técnicas bayesianas y por máxima verosimilitud, que en algunos campos de investigación están sustituyendo a los métodos de la llamada estadística frecuentista.

### REFERENCIAS BIBLIOGRÁFICAS

Anderson, M.J. & Legendre, P. 1999. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J. Statist. Comput. Simulation* 62: 271-303.

Avanza, M., Mazza, S., Martínez, G. & Giménez, L. 2003. Aplicación de transformaciones para el cumplimiento de los supuestos de normalidad y homocedasticidad a concentraciones foliares de N, P y K en mandarino. *Agrotécnica* 11: 18-23.

Bandera, E. 2018. Estudio de la variabilidad genética y de la asociación entre caracteres cuantitativos en familias de hermanos completos de guayabo (*Psidium guajava* L.). Tesis de Maestría. Facultad de Biología, Universidad de la Habana.

Cheng, J.G. 2000. Outlier management in intelligent data analysis. Tesis de Doctorado. Birkbeck College, University of London, England.

De Farias Neto, J.T. & Resende, M.D.V. 2001. Aplicação da metodologia de modelos mistos (REML/BLUP) na estimação de componentes de variância e predição de valores genéticos em pupunheira (*Bactris gasipaes*). *Rev. Bras. Frutic.* 23(2): 320-324.

Di Rienzo, J.A., Casanoves, F., González, L.A., Tablado, E.M., Díaz, M.P., Robledo, C.W. & Balzarini, M.G. 2005. Estadística para las ciencias agropecuarias. Sexta Edición. Editorial Brujas, Córdoba, Argentina.

Efron, B. & Tibshirani, R.J. 1993. An introduction to the bootstrap. Monographs on Statistics and Applied Probability, No. 57. Chapman & Hall, London, England.

Fagerland, M.W. & Sandvik, L. 2009. The Wilcoxon-Mann-Whitney test under scrutiny. *Stat. Med.* 28: 1487-1497.

Fagerland, M.W. 2012. t-tests, non-parametric tests, and large studies – a paradox of statistical practice? *BMC Medical Research Methodology* 12: 78.

Feng, Ch., Wang, H., Lu, N., Chen, T., He, H., Lu, Y. & Tu, X.M. 2014. Log-transformations and its implications for data analysis. *Shanghai Arch. Psychiatry* 26: 105-109.

Garson, G.D. 2012. Testing statistical assumptions. Statistical Associates Publishing. Asheboro, NC, USA.

Gotelli, N.J. & Graves, G.R. 1996. Null models in ecology. Smithsonian Institution Press, Washington DC, USA.

Gotelli, N.J. 2001. Research frontiers in null model analysis. *Global Ecology & Biogeography* 10: 337-343.

Grissom, R.J. 2000. Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology* 68: 155-165.

Herrera, M., Guerra, C.W., Sarduy, L., García, Y. & Martínez, C.E. 2012. Diferentes métodos estadísticos para el análisis de variables discretas. Una aplicación en las ciencias agrícolas y técnicas. *Rev. Cie. Téc. Agr.* 21(1):58-62.

Howell, D.C. 2007. Statistical methods for psychology. Seventh Edition. University of Vermont. Belmont, CA, USA.

Judd, C.M., McClelland, C.H. & Ryan, C.S. 2009. Data analysis: a model-comparison approach. Second Edition. Routledge, New York, NY, USA.

Krutchkoff, R.G. 1998. One-way fixed effects analysis of variance when the error variances may be unequal. *J. Statist. Comput. Simula.* 30: 259-271.

Littell, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. 2006. SAS for Mixed Models. 2nd Edition, SAS Institute Inc., Cary, NC, USA.

Manikandan, S. 2010. Data transformation. *J. Pharmacol. Pharmacother.* 1(2): 126-27.

Manly, B.F.J. 2007. Randomization, Bootstrap and Monte Carlo Methods in Biology. Third Edition. Chapman and Hall, Boca Raton FL, USA.

- McDonald, J.H. 2014. Handbook of Biological Statistics. Sparky House Publishing, Baltimore, Maryland, USA.
- Miller, R.G. 1997. Beyond ANOVA: Basics of applied statistics. Chapman & Hall, New York, USA.
- Murphy, T. & Lau, A.T. 2008. Manejo de valores atípicos. ¿Cómo se evalúa un valor aberrante o inconsistente único? ASTM Standardization News.
- Oehlert, G.W. 2012. A few words about REML. University of Minnesota, Ford Hall, SE, USA
- Orr, J.M., Sackett, P.R. & DuBois, C.L. 1991. Outlier detection and treatment in I/O psychology: A survey of research beliefs and an empirical illustration. *Personnel Psychology* 44: 473-486.
- Osborne, J.W. 2002. Notes on the use of data transformations. *Practical Assessment, Research & Evaluation* 8(6): 1-9.
- Osborne, J.W. 2008. Best Practices in Data Transformation: The overlooked effect of minimum values. pp: 197-204. En: Osborne, J.W. (ed.) Best Practices in Quantitative Methods. Sage Publishing, Thousand Oaks, CA, USA.
- Osborne, J.W. 2010. Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation* 15(12): 1-9.
- Piepho, H.P., Möhring, J., Melchinger, A.E. & Büchse, A. 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161: 209-228.
- SAS Institute Inc. 2012. SAS/IML 9.3 User's Guide. SAS Institute Inc., Cary, NC. URL <http://www.sas.com/>
- Stroup, W. 2012. Generalized linear mixed models: Modern concepts, methods and applications. En: Texts in Statistical Science. Chapman & Hall, London, England.
- Stroup, W.W. 2015. Rethinking the Analysis of Non-Normal Data in Plant and Soil Science. *Agron. J.* 107: 811-827.
- Tabachnick, B.G. & Fidell, L.S. 2007. Using multivariate statistics. Allyn and Bacon, Boston, USA.
- Wang, T., He, P., Ahn, K.W., Wang, X., Ghosh, S. & Laud, P. 2015. A re-formulation of generalized linear mixed models to fit family data in genetic association studies. *Frontiers in Genetics* 6: 120.
- Whitlock, M.C. & Schluter, D. 2009. The Analysis of Biological Data. Roberts and Company Publishers, Greenwood Village, Colorado, USA.
- Zar, J.H. 2010. Biostatistical Analysis. 5th ed. Pearson Prentice Hall, New Jersey, USA.
- Zhou, M.M. & Kimbeng, C.A. 2010. Multivariate repeated measures: A statistical approach for analyzing data derived from sugarcane breeding variety trials. *Proc. S. Afr. Sug. Technol. Ass.* 83: 92-105.
- Zimmerman, D.W. 2000. Statistical significance levels of non-parametric tests biased by heterogeneous variances of treatment groups. *J. Gen. Psychol.* 17: 354-364.