
1. TRANSFORMACIONES Y ESTANDARIZACIONES

INTRODUCCIÓN

Ejemplo 1. TRANSFORMACIONES GENERALES

1. Criterios generales en las transformaciones

Ejemplo 2. TRANSFORMACIONES Y ESTANDARIZACIONES

2. Regla de abultamiento de Mosteller y Tukey (1977).

Ejemplo 3. REGLA DE ABULTAMIENTO

3. Ley de potencias de Taylor (Taylor 1961)

Ejemplo 4. LEY DE POTENCIAS DE TAYLOR

4. Transformación poder de Box-Cox (1964).

Ejemplo 5. LEY DE PODER DE BOX-COX.

Ejemplo 6. LEY DE PODER DE BOX-COX.

EJERCICIOS PROPUESTOS

BIBLIOGRAFÍA

Javier Rodríguez-Barrios

INTRODUCCIÓN

Los estudios tradicionales en ciencias biológicas, que se apoyan en el método científico, pueden considerar pruebas estadísticas que soporten a las pruebas de hipótesis diseñadas para dar respuesta a preguntas de investigación. En este caso, es importante contar con bases de datos de buena calidad, que permitan un buen análisis de la información requerida. Este análisis puede realizarse en principio, por medio de la exploración gráfica y numérica de los datos. Sin embargo, en determinados casos, el análisis exploratorio de los datos, con las diferentes opciones de visualización gráfica (ver capítulo de exploración en R), requiere de mejoras en la resolución de los datos, para que estos presenten patrones y simetrías definidas, que se ajusten mejor a los diseños estadísticos que permitan dar respuesta a las hipótesis realizadas. Para este caso, las transformaciones y estandarizaciones de los datos, pueden ser una buena alternativa de ajuste.

Las transformaciones en un contexto estadístico se representan por expresiones matemáticas que se pueden aplicar a los datos organizados en filas (observaciones) o a las columnas (variables), en una matriz que contiene a la información tabulada. En el contexto biológico, las transformaciones son herramientas que permiten ajustar los datos a distribuciones esperadas, de forma similar, las estandarizaciones permiten hacer más comparativas a las variables en el análisis, especialmente aquellas que presenten escalas o magnitudes disímiles. De acuerdo a McCune y Grace (2002), Las transformaciones permiten encontrar la diferencia entre lo anhelado y lo actual, la oscuridad y la claridad. Una buena transformación depende del entendimiento detallado de los datos y de la claridad de lo que se busca. La selección de la mejor transformación por lo general es compleja y se escoge por experiencia del investigador. Diferentes estrategias son disponibles para escoger la mejor transformación. (1) la primera opción es mediante pruebas de ensayo y error, (2) el uso de figuras exploratorias también suele ser importante, como el caso de figuras de dispersión de puntos, cajas y graficas de pares, entre otras.

De todas formas es importante revisar que los datos hayan sido bien tabulados y el instrumento de medición de la variable sea el apropiado y este calibrado. Los valores atípicos deben evaluarse en figuras exploratorias para considerar si es necesario excluirllos del análisis y permitir una mejor distribución de los datos o si su significado es importante para no excluirllos. El presente capítulo, intenta resumir diferentes tipos de estrategias utilizadas en el ámbito biológico, para la escogencia de la transformación más adecuada cuando esta sea requerida.

Las transformaciones, normalmente están precedidas de figuras exploratorias, dependiendo de la naturaleza de los datos, estas figuras pueden representarse por histogramas (figura 1). En estudios biológicos, es muy frecuente encontrar que los datos se concentran o presentan mayor frecuencia de valores bajos, generando asimetrías hacia la derecha (figura 1a), lo cual puede ser un indicador de ausencia de simetría de distribución normal. Para estos casos, suelen realizarse transformaciones sencillas y rápidas como las logarítmicas (figura 1b) o de raíces cuadrada y cuarta (figuras 1 c, d).

Ejemplo 1. TRANSFORMACIONES GENERALES

El presente ejercicio consiste en la realización de histogramas de frecuencia, de una base de datos de proyecto desarrollado en el piedemonte occidental de Nariño, con el propósito de establecer si existe o no relación entre las medidas morfológicas y la dieta o hábito alimenticio de las aves registradas. Para esto se registraron las medidas estándar que toman los ornitólogos en museo y por literatura se asignaron los hábitos de cada una de las especies. (Fuente: Cepeda Belisario (2011) – En prensa). Los archivos que se utilizarán son: **Aves.csv** y **Aves.r**.

Lectura de la base de datos.

```
aves<-read.csv2("Aves.csv",row.names=1)
```

Paquete requerido

```
library(lattice)
```

Realización de los histogramas

```
histogram(~Longitud.total, data=aves, nint= 7,
```

```
  xlab = "Longitud Total de Aves (mm)", ylab = "Frecuencias")
```

```
histogram(~log10(Longitud.total), data=aves, nint= 7,
```

```
  xlab=expression(log[10]*(Longitud_Total)),ylab="Frecuencias")
```

```
histogram(~sqrt(Longitud.total), data=aves, nint= 7,
```

```
  xlab = "(Longitud Total)^1/2", ylab = "Frecuencias")
```

```
histogram(~sqrt(sqrt(Longitud.total)), data=aves, nint= 7,
```

```
  xlab = "Raíz Cuarta (Longitud Total de Aves)", ylab = "Frecuencias")
```

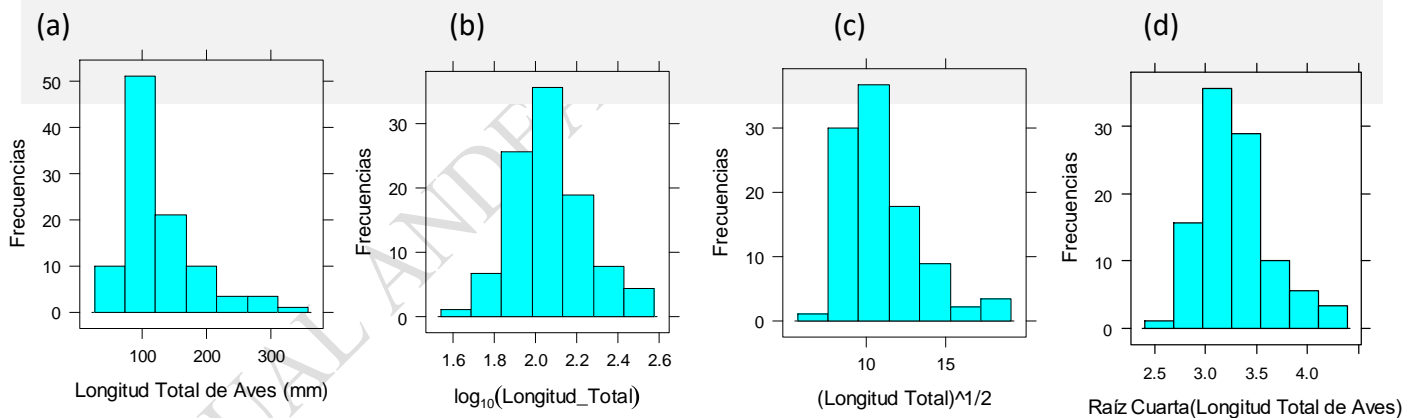


Figura 1. Transformaciones de la variable longitud total de aves en el Departamento de Nariño (tomado de Cepeda 2010). (a) Medidas de la variable en milímetros, (b) medidas en logaritmo base 10, (c) medidas bajo la transformación raíz cuadrada y (d) doble raíz cuadrada (raíz cuarta).

En este sentido, las transformaciones sencillas, como las realizadas con logaritmo base 10 (\log_{10}) presentan varios puntos a favor. (1) Los datos originales que para el caso del ejemplo 1, son tomados en milímetros pasan a ser medidos en escala logarítmica; (2) los datos se transforman para volverlos más normales (normalizarlos); (3) esta transformación permite realizar inferencia de hipótesis mediante pruebas estadísticas paramétricas (más robustos que las técnicas paramétricas).

1. Criterios generales en las transformaciones

Las transformaciones suelen hacerse en datos que no cumplen con el requisito de una distribución normal, debido en su mayoría a la presencia de datos atípicos, o por la falta de homogeneidad en las varianzas de las variables (heterogeneidad). Dichas transformaciones se realizan como criterio requerido en técnicas estadísticas que exigen los patrones mencionados (normalidad u homogeneidad). De acuerdo a Borcard et al. (2018), se puede resumir las situaciones en las cuales debe aplicarse una transformación.

- **Volver las unidades comparativas.** Para el caso de variables que se presenten en diferentes unidades de medida y que se quieran relacionar o incluir en un mismo análisis, como puede ser el caso de variables físico-químicas (estandarizaciones en rangos que pueden ser de -1 a 1 o de 0 a 1, con lo cual se pierde la dimensionalidad original de las variables).
- **Volver las variables simétricas.** Para el caso de variables que requieren presentar una distribución normal, se realizan transformaciones de estos descriptores, para estabilizar sus variaciones (transformaciones, raíces cuadrada, cuarta, logaritmos, inversos, recíprocos, angulares, etc.).
- **Linealizar las relaciones en las variables.** Para el caso de variables que requieran ser linealizadas, debido a requerimientos de algunas técnicas estadísticas (transformaciones logarítmicas en variables que presenten tendencias exponenciales).
- **Variables que enmascaran al resto.** Ocurre frecuentemente en variables como por ejemplo, las abundancias de algunas especies, estimada por conteos, densidades, coberturas, biomasa, entre otros, allí aparecen especies dominantes que enmascaran los patrones del resto de organismos (transformaciones raíces cuadrada o cuarta, logaritmos, sumando 1 a todos los datos). Hay casos extremos en los que las abundancias deben transformarse a datos binarios, de presencias (1) y ausencias (0).

Según McCune y Grace (2002) Las transformaciones de datos, especialmente ecológicos, se pueden agrupar en tres grandes grupos (1) monotónicas, (2) suavizadas y (3) relativizadas.

1.1 Transformaciones monotónicas o “simples”. Transformaciones que se aplican solo a alguna(s) variable(s) u observaciones (fila o columna, respectivamente).

x_{ij} = el valor original en la fila i y la columna j de la matriz de datos

b_{ij} = el valor ajustado o transformado que reemplaza x_{ij} .

Transformaciones de poder o potencia. Aplica principalmente a datos discretos de tipo conteo (ej. Abundancias de individuos), Se transforman los datos con diferentes parámetros o exponentes (p).

$$b_{ij} = x_{ij}^p$$

Si, $p=0.5$, la transformación será raíz cuadrada. Si, $p=0.25$, la transformación será raíz cuarta. Para el caso de la raíz cuadrada, el efecto de su transformación es similar a la transformación logarítmica, pero es más fuerte.

Transformaciones superiores a la cúbica, aproximan los datos a presencia – ausencia (binarios), para este caso, los valores cercanos a cero se transforman en uno y los ceros quedan iguales.

Transformaciones logarítmicas. Permiten ajustar la distribución de datos que presentan simetrías hacia la derecha, ocasionados por la concentración en valores bajos (figura 1). Estas transformaciones son útiles variables que relacionan longitudes, pesos, volúmenes, áreas, etc. También son útiles cuando la variación (varianza) de los datos es alta, como en conteos de abundancia o mediciones de biomasa. También aplica para una gran variedad de variables ambientales y de hábitas.

$$b_{ij} = \log(x_{ij})$$

Para el caso de abundancias, que suelen presentar ceros, se suele adicionar 1 a la transformación, debido a que el logaritmo de cero es indeterminado.

$$b_{ij} = \log(x_{ij} + 1)$$

La adición de 1, no es recomendable si se tienen fracciones pequeñas, debido a que afecta la diferencia de magnitudes, como puede ocurrir al adicionar 1 a valores de 0,02 y 0,002, en este caso, la transformación no permitirá visualizar la proporcionalidad de las diferencias, debido a que al sumar 1, los logaritmos de 1,02 y de 1,003 representan valores similares.

Transformación arcoseno. Conocida también como seno⁻¹ o el inverso del seno (1/seno), esta transformación es apropiada para variables cuantificadas en porcentajes o proporciones. Para este tipo de transformaciones, los datos deben oscilar entre cero y uno. Esta transformación suele ser inversa a la logarítmica, pues permite corregir distribuciones que presentan asimetría a la izquierda, que obedecen a una mayor frecuencia de valores altos. En este sentido, suelen ser útil para valores con sesgos negativos.

$$b_{ij} = \frac{2}{\pi} \cdot \arcsen(x_{ij})$$

Transformación raíz cuadrada del arcoseno. Útil para datos de proporciones, que presentan mayor frecuencia de valores bajos y altos, con pocos datos de magnitudes intermedias.

$$b_{ij} = \frac{2}{\pi} \cdot \arcsen(\sqrt{x_{ij}})$$

Transformación logit. Útil para datos de proporciones, pero en valores de 1 y 0, el logit es indefinido.

$$b_{ij} = \ln(x_{ij}/(1 - x_{ij}))$$

1.2 Suavizamiento de Beals (beals smoothing). Consiste en reemplazar los datos por la probabilidad de que variables como abundancias de especies (columnas), ocurran en sitios en particular (fila), en base a las ocurrencias de las especies, objetivo, con las que realmente están en el sitio. No es apropiada para variables continuas o que tengan decimales, debido a que limita su proporcionalidad.

Busca corregir el problema de los ceros en las variables (ej. Ausencias de especies en algunas localidades), dando más valor o “favorabilidad” a las variables (i) de cada muestra, con más valores de 1 (ej. Presencias) en el conjunto de datos totales, usando las proporciones de ocurrencias conjuntas entre las variables con valores en la muestra (valor 1) y la variable i. El método puede ser útil para identificar la pertenencia de especies a su grupo (Ewald 2002).

$$b_{ij} = \frac{1}{S_i} \cdot \sum_k \left(\frac{M_{jk} \cdot I_{ik}}{N_k} \right)$$

Corresponde a la probabilidad estimada (b_{ij}) que la especie j , ocurra en el sitio i . Donde, S_i es el número de especies en una muestra unitaria o sitio i , M_{jk} es el número de ocurrencias de las especies j y k . N_k es el número de ocurrencias de la especie k . I es la incidencia (0 a 1) de las especies (termino usualmente omitido en la ecuación pero necesario).

1.3 Relativizaciones o estandarizaciones. Permiten re-escalar filas o columnas individuales, en relación a algún criterio definido (valor máximo, total, promedio, etc.). Es útil cuando existe disimilitud marcada en las magnitudes de las variables o de las observaciones. El coeficiente de variación porcentual (%CV), permite indicar el grado de variación en los totales de las variables. Si el %CV es menor de 50, la variabilidad se considera baja. Para un %CV entre 50 y 100, la variabilidad es moderada. Si el %CV mayores a 300, indica variaciones muy altas.

Las relativizaciones permiten estandarizar o poner en igualdad de condiciones a variables que hayan sido medidas en escalas diferentes (ej. Físico-químicas o abundancias, combinadas con coberturas y área basal). Tienen un efecto importante en variables que enmascaran a otras por su elevada magnitud. Para estos casos, el desempeño de relativizaciones como las *correlaciones* o el procedimiento de restar a cada dato de la variable por su promedio y dividir por su desviación estándar, es más eficiente que otras transformaciones.

Relativizar por el valor máximo. Permite volver comparativas a valores bajos y altos de las variables (i) en cada muestra u observación (j).

$$b_{ij} = x_{ij}/x_{j\text{máximo}}$$

La salida variará de 0 a 1, los valores más grandes se escalarán a 1. Se convierten valores en un valor relativo (igual al valor máximo). Se utiliza esta transformación cuando hay diferencias en los totales de las variables.

Relativizar por el valor promedio. Al restar cada valor por la media, produce datos positivos y negativos.

$$b_{ij} = x_{ij} - \bar{x}_j$$

Relativizar por la desviación estándar. Conocido como “puntajes z ”, permite que las variables presenten media de cero y varianza de 1. También genera datos positivos y negativos. S_j es la variable i .

$$b_{ij} = (x_{ij} - \bar{x}_j)/S_i$$

Se puede aplicar a cualquier variable cuantitativa, convierte valores en puntajes z (media = 0, varianza = 1), volviendo más comparativas a las variables, se recomienda para variables en diferentes escalas o unidades de medida. Para hacer esta transformación, se resta cada elemento por la media de la columna y luego se divide por la desviación estándar de la columna. El resultado es un valor que representa el número de desviaciones estándar de del promedio de la variable.

Relativizar datos binarios con respecto a su mediana. A u elemento se le asigna cero, si su valor es menor o igual a la mediana de su fila o de su columna y uno si está por encima de su mediana. Lo anterior permite hacer comparativa de variables que se encuentran en rangos.

$$b_{ij} = 1, \text{ si } x_{ij} > \text{mediana}, \quad b_{ij} = 0, \text{ si } x_{ij} \leq \text{mediana}$$

Estandarización por Normalización. Útil cuando las observaciones (filas) presentan mucha dispersión o varianza, en este caso, a cada dato se le resta su valor máximo y se le divide la diferencia de su máximo y mínimo. Se puede aplicar a cualquier variable cuantitativa, los valores de b_{ij} oscilarán entre 0 y 1.

$$b_{ij} = x_{ij} - x_{j\text{min}}/(x_{j\text{max}} - x_{j\text{min}})$$

Donde, $x_{j\text{max}}$ y $x_{j\text{min}}$ son los valores máximos y mínimos de las observaciones o muestras (filas)

Estandarización de Hellinger. Similar a la anterior, es utilizada para estandarizar filas u observaciones. Cada dato se divide por su suma de filas. Después se calcula la raíz cuadrada de cada elemento. Se puede aplicar a cualquier variable cuantitativa, los valores de b_{ij} serán superiores a 1.

$$b_{ij} = \sqrt{(x_{ij}/(x_{j\text{total}}))}$$

La distancia de Hellinger es recomendada por Legendre y Gallagher (2001), debido a que permite Linealizar datos (observaciones o filas) con mucha dispensación y valores de ceros, como los valores de abundancias de especies. Esta distancia es conocida como la distancia Euclídea al cuadrado.

Estandarización de Chi Cuadrado. A cada elemento se divide por el máximo de su columna y luego se divide por el total de la fila. Brinda igual ponderación a variables o muestras. Se puede aplicar a cualquier variable cuantitativa superior a cero, los valores de b_{ij} oscilan entre 0 y 1.

1.4 Paquetes de R que integran transformaciones. A continuación se describen algunos paquetes, documentados por Borcard et al. (2018), que incorporan diferentes tipos de transformaciones y estandarizaciones generales, a variables a y observaciones. El paquete *vegan* es comúnmente utilizado en estudios ecológicos, debido a que incorpora herramientas univariadas y multivariadas en análisis estadísticos.

Estandarizaciones en *vegan*. El paquete “*vegan*” contiene el comando “*decostand*”, que cuenta con diferentes opciones, para realizar transformaciones y estandarizaciones de variables (columnas) y de observaciones (filas). Incluye transformaciones “simples” como la logarítmica, la raíz cuadrada, “estandarizaciones generales” como el máximo, el mínimo, “estandarizaciones por filas” como la Hellinger o la normalización, así como “dobles estandarizaciones”, como la Chi Cuadrado o la Wisconsin.

Estandarizaciones con el paquete *labds*. El paquete “*vegtrans*” aplica el comando “*vegtrans*” en matrices de datos que combinan diferentes tipos de variables en un mismo estudio, como los de vegetación, en los que se tiene información de abundancias, coberturas, abundancias y coberturas a la vez (escala de dominancia de Braun-Blanquet). Para estos casos se transforman las escalas, atribuyendo valores de acuerdo a los datos disponibles.

Tabla 1. Transformaciones propuestas por McCune y Grace (2002) y comandos en R para ejecutarlas. Donde x_{ij} es la variable u observación sin transformar y b_{ij} es la variable u observación transformada.

Transformaciones	Valores de x_{ij}	Valores de b_{ij}	Comando en R
1. Monotónicas			
$x^{1/2}$ (poder)	sin negativos	sin negativos	<code>sqrt(x)</code>
$x^{1/4}$ (poder)	sin negativos	sin negativos	<code>sqrt(sqrt(x))</code>
Log (x)	positivos	todos	<code>log10(x)</code>
$(2/\pi) \cdot \arccos(x^{1/2})$	0 a 1	0 a 1	<code>asin(sqrt(x))*180/pi</code>
2. Suavizamientos de Belt			
Suavizamiento de beals	Conteos	0 a 1	<code>beals(x)</code>
3. Relativizaciones			
Máximos por filas	no negativo	0 a 1	<code>decostand(x, "max", 2)</code>
Máximos por columnas	no negativo	0 a 1	<code>decostand(x, "max", 1)</code>
Por la desviaciones estándar	todos	-10 a 10	<code>scale (x)</code>
Binario por el promedio	todos	0 o 1	--
Binario por mediana	todos	0 o 1	--
Normalización	todos	0 o 1	<code>decostand(x, "normalize")</code>
Hellinger	todos	0 o 1	<code>decostand(x, "hell")</code>
Chi Cuadrado	todos	0 o 1	<code>decostand(x, "chi.square")</code>
Wisconsin	todos	0 o 1	<code>wisconsin(x)</code>

x es la base de datos, solo con las variables cuantitativas.

MANUAL ANDEA - JAVIER RODRÍGUEZ B.

Ejemplo 2. TRANSFORMACIONES Y ESTANDARIZACIONES

El presente ejercicio consiste en realizar diferentes tipos de transformaciones y estandarizaciones generales, basado en datos hipotéticos de cinco órdenes de insectos acuáticos, evaluados en diez localidades.

```
# Lectura de la base de datos
datos<-read.csv2("Insectos.csv",row.names=1)
datos<-datos[,2:6]
head(datos)
# Paquete requerido
library(vegan)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	3	4	4	6	1
S2	5	1	1	7	3
S3	6	2	0	2	6
S4	1	1	1	0	3
S5	4	7	3	6	2
S6	2	2	5	1	0

1.) Transformaciones monotónicas

Transformación raíz cuadrada

```
datos.r= sqrt(datos)
head(datos.r)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	1.732051	2.000000	2.000000	2.449490	1.000000
S2	2.236068	1.000000	1.000000	2.645751	1.732051
S3	2.449490	1.414214	0.000000	1.414214	2.449490
S4	1.000000	1.000000	1.000000	0.000000	1.732051
S5	2.000000	2.645751	1.732051	2.449490	1.414214
S6	1.414214	1.414214	2.236068	1.000000	0.000000

Transformación logarítmica

```
datos.log= log10(datos)
head(datos.r)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	1.732051	2.000000	2.000000	2.449490	1.000000
S2	2.236068	1.000000	1.000000	2.645751	1.732051
S3	2.449490	1.414214	0.000000	1.414214	2.449490
S4	1.000000	1.000000	1.000000	0.000000	1.732051
S5	2.000000	2.645751	1.732051	2.449490	1.414214
S6	1.414214	1.414214	2.236068	1.000000	0.000000

Frecuencia relativa, por especies

```
datos.fr= prop.table(datos, margin=NULL)
```

Frecuencia relativa de sitios

```
datos.s.fr <- decostand(datos, "total", MARGIN = 2)
```

Transformación arcoseno

```
daros.arc1= asin(sqrt(datos.fr))*180/pi
head(daros.arc1)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	8.185225	9.462322	9.462322	11.615728	4.715004
S2	10.591391	4.715004	4.715004	12.561030	8.185225
S3	11.615728	6.675592	0.000000	6.675592	11.615728
S4	4.715004	4.715004	4.715004	0.000000	8.185225
S5	9.462322	12.561030	8.185225	11.615728	6.675592
S6	6.675592	6.675592	10.591391	4.715004	0.000000

Proporciones para transformación arcoseno

```
datos.prop <- datos / apply(datos,1,sum)
```

Transformación arcoseno

```
daros.arc2= asin(sqrt(datos.prop))*180/pi
head(daros.arc2)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	8.185225	9.462322	9.462322	11.615728	4.715004
S2	10.591391	4.715004	4.715004	12.561030	8.185225
S3	11.615728	6.675592	0.000000	6.675592	11.615728
S4	4.715004	4.715004	4.715004	0.000000	8.185225
S5	9.462322	12.561030	8.185225	11.615728	6.675592
S6	6.675592	6.675592	10.591391	4.715004	0.000000

2.) Suavizado (Beals)

Lectura de datos

```
datos.belt <- beals(datos)
head(datos.belt)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.8266667	1	0.9105556	0.9105556	0.9105556
S2	0.8266667	1	0.9105556	0.9105556	0.9105556
S3	0.8388889	1	0.8881944	0.9159722	0.9159722
S4	0.8388889	1	0.9159722	0.8881944	0.9159722
S5	0.8266667	1	0.9105556	0.9105556	0.9105556
S6	0.8388889	1	0.9159722	0.9159722	0.8881944

Ejemplo 1. TRANSFORMACIONES Y ESTANDARIZACIONES (Cont.)

3.) Estandarizaciones

3.1 Máximo por columnas (variables)

Máximo para especies

```
datos.c <- decostand(datos, "max", 1)
head(datos.c)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.4285714	0.5714286	0.8	0.8571429	0.1666667
S2	0.7142857	0.1428571	0.2	1.0000000	0.5000000
S3	0.8571429	0.2857143	0.0	0.2857143	1.0000000
S4	0.1428571	0.1428571	0.2	0.0000000	0.5000000
S5	0.5714286	1.0000000	0.6	0.8571429	0.3333333
S6	0.2857143	0.2857143	1.0	0.1428571	0.0000000

Relación media y desviación - Puntaje Z

```
datos.s <- scale(datos)
head(datos.s)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.0000000	0.2587318	0.6404113	1.2026756	-0.8183932
S2	0.8164966	-1.0349272	-0.9606169	1.6035675	0.2045983
S3	1.2247449	-0.6037076	-1.4942929	-0.4008919	1.7390856
S4	-0.8164966	-1.0349272	-0.9606169	-1.2026756	0.2045983
S5	0.4082483	1.5523909	0.1067352	1.2026756	-0.3068975
S6	-0.4082483	-0.6037076	1.1740873	-0.8017837	-1.3298890

3.2 Estandarización para filas (sitios)

Máximo para sitios

```
datos.f = decostand(datos, "max", 2)
head(datos.f)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.4285714	0.5714286	0.8	0.8571429	0.1666667
S2	0.7142857	0.1428571	0.2	1.0000000	0.5000000
S3	0.8571429	0.2857143	0.0	0.2857143	1.0000000
S4	0.1428571	0.1428571	0.2	0.0000000	0.5000000
S5	0.5714286	1.0000000	0.6	0.8571429	0.3333333
S6	0.2857143	0.2857143	1.0	0.1428571	0.0000000

Normalización

```
datos.norm <- decostand(datos, "normalize")
head(datos.norm)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.3396831	0.4529108	0.4529108	0.6793662	0.1132277
S2	0.5423261	0.1084652	0.1084652	0.7592566	0.3253957
S3	0.6708204	0.2236068	0.0000000	0.2236068	0.6708204
S4	0.2886751	0.2886751	0.2886751	0.0000000	0.8660254
S5	0.3746343	0.6556101	0.2809757	0.5619515	0.1873172
S6	0.3429972	0.3429972	0.8574929	0.1714986	0.0000000

Hellinger

```
datos.hell <- sqrt(datos / apply(datos, 1, sum))
head(datos.hell)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.4082483	0.4714045	0.4714045	0.5773503	0.2357023
S2	0.5423261	0.2425356	0.2425356	0.6416889	0.4200840
S3	0.6123724	0.3535534	0.0000000	0.3535534	0.6123724
S4	0.4082483	0.4082483	0.4082483	0.0000000	0.7071068
S5	0.4264014	0.5640761	0.3692745	0.5222330	0.3015113
S6	0.4472136	0.4472136	0.7071068	0.3162278	0.0000000

3.2 Doble estandarización (filas y columnas)

Chi Cuadrado

```
datos.chi <- decostand(datos, "chi.square")
head(datos.chi)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.3701851	0.4636379	0.5109040	0.7403703	0.1325475
S2	0.6532679	0.1227277	0.1352393	0.9145750	0.4210334
S3	0.8329166	0.2607963	0.0000000	0.2776389	0.8946959
S4	0.3701851	0.3477284	0.3831780	0.0000000	1.1929279
S5	0.4038383	0.6638451	0.3135093	0.6057575	0.2168960
S6	0.4442222	0.4172741	1.1495341	0.2221111	0.0000000

Wisconsin

```
datos.wis <- wisconsin(datos)
head(datos.wis)
```

	Coleop	Dipter	Himeno	Hemip	Lepid
S1	0.1517707	0.20236088	0.28330523	0.30354132	0.05902192
S2	0.2793296	0.05586592	0.07821229	0.39106145	0.19553073
S3	0.3529412	0.11764706	0.00000000	0.11764706	0.41176471
S4	0.1449275	0.14492754	0.20289855	0.00000000	0.50724638
S5	0.1699717	0.29745042	0.17847025	0.25495751	0.09915014
S6	0.1666667	0.16666667	0.58333333	0.08333333	0.00000000

Ejemplo 1. TRANSFORMACIONES Y ESTANDARIZACIONES (Cont.)

Figuras de cajas con algunas transformaciones para los coleópteros

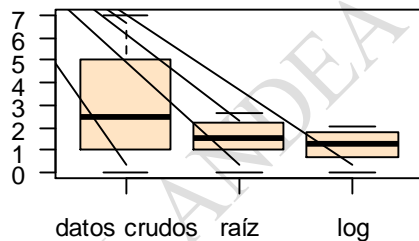
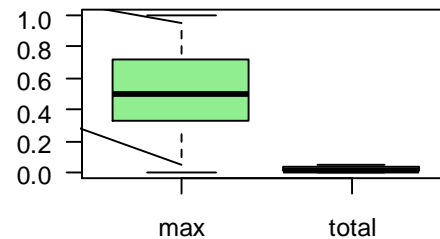
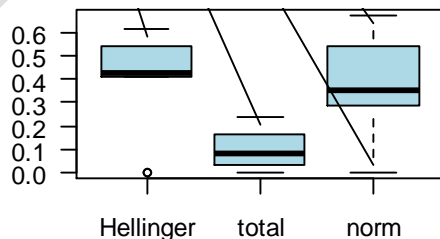
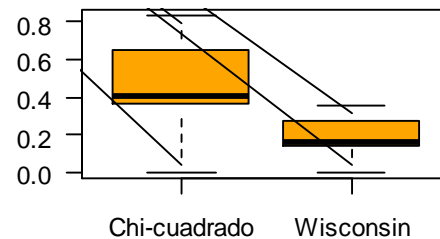
```
par(mfrow = c(2,2))
boxplot(datos$Coleop, sqrt(datos$Coleop), log1p(datos$Coleop), las = 1,
        main = "Transformaciones simples", names = c("datos crudos", "raíz", "log"), col = "bisque")

boxplot(datos.c$Coleop, datos.fr$Coleop, las = 1,
        main = "Estandarizaciones para especies", names = c("max", "total"), col = "lightgreen")

boxplot(datos.hell$Coleop, datos.s.fr$Coleop, datos.norm$Coleop, las = 1,
        main = "Estandarizaciones para sitios", names = c("Hellinger", "total", "norm"), col = "lightblue")

boxplot(datos.chi$Coleop, datos.wis$Coleop, las = 1,
        main = "Doble estandarización", names = c("Chi-cuadrado", "Wisconsin"), col = "orange")

# Se cierra el panel diseñado
par(mfrow = c(1,1))
```

Transformaciones simples**Estandarizaciones para especies****Estandarizaciones para sitios****Doble estandarización**

2. Regla de abultamiento de Mosteller y Tukey (1977).

Ejemplo 3. REGLA DE ABULTAMIENTO

3. Ley de potencias de Taylor (Taylor 1961)

Ejemplo 4. LEY DE POTENCIAS DE TAYLOR

4. Transformación poder de Box-Cox (1964).

Ejemplo 5. LEY DE PODER DE BOX-COX.

Ejemplo 6. LEY DE PODER DE BOX-COX.

EJERCICIOS PROPUESTOS

BIBLIOGRAFÍA

Nota: Se requiere contar con la versión completa del libro Análisis de Datos Ambientales y Ecológicos - ANDEA