

Universidad Jorge Tadeo Lozano

Estadística para la Biología y Ecología

Estadística descriptiva, univariada y multivariada

Andrés Pulido Hernández

<http://www.geocities.com/biologiamar>



Universidad de Bogotá
JORGE TADEO LOZANO

09

Tabla de contenido

1. INTRODUCCIÓN	3
1.1. Tipos de datos	4
1.2. Tipos de variables.....	5
1.3. Tipos de pruebas estadísticas univariadas.....	5
2. ESTADÍSTICA DESCRIPTIVA.....	6
2.1. Medidas de dispersión y variabilidad.....	7
3. MÉTODOS UNIVARIADOS.....	7
3.1. Distribución normal.....	7
3.1.1. Planteamiento de hipótesis.....	8
3.1.2. Tipos de pruebas estadísticas	9
3.1.3. Propiedades de la distribución normal	10
4. EJECUCIÓN DE LAS PRUEBAS ESTADÍSTICAS UNIVARIADAS	11
5. MÉTODOS MULTIVARIADOS	19
5.1. Índices de Diversidad	19
5.1.1. Instrucciones del manejo del programa PRIMER v5.0	19
5.1.2. Análisis de matriz de diversidad.....	22
5.2. Análisis de clasificación	23
5.2.1. Cluster	23
5.2.2. MDS.....	24
6. BIBLIOGRAFÍA RECOMENDADA.....	25
7. BIBLIOGRAFÍA CITADA.....	26

1. INTRODUCCIÓN

La estadística es una disciplina que se encarga de coleccionar, organizar, resumir, analizar e interpretar datos con el propósito de obtener inferencias objetivas y reales a partir de un conjunto de datos. Es aplicada en la biología y la ecología donde cambia los datos descriptivos a cuantitativos para formular hipótesis y someterlas a pruebas estadísticas. Es muy importante porque en algunos casos permite saber cuantitativamente el grado de certeza de las conclusiones a las que se llega (Ramírez, 2005; Sanjuan, 2006).

En ésta guía, se tratan tres tipos básicos de métodos, una es la **estadística descriptiva** la cual no dice mucho, pero es muy útil para empezar cualquier proyecto. Los **métodos univariados** (de un solo factor > véase *abajo* la definición) aplicado a análisis de poblaciones, en los cuales se deben proponer una hipótesis nula y una alterna; es importante notar que *no siempre* el rechazo de la hipótesis nula no comprueba la hipótesis alterna. Luego se debe establecer una probabilidad límite antes de realizar la prueba (para este caso 0,05 > 95% de confianza) y las tablas con las que se confrontan los resultados se organizan con el valor de probabilidad y el tamaño de la muestra (Sanjuan, 2006). El tercer **método es el multivariado**, aplicado a análisis de comunidades, donde se pueden comparar diferentes estratos definidos según el investigador según sea el caso, por ejemplo: entre transectos, entre parcelas, entre métodos de captura, entre ecosistemas, etc.

NOTA: La prueba estadística determina la significancia estadística, pero no la biológica la cual debe ser deducida por el investigador.

La **estadística** si se usa de un modo adecuado, es muy útil y potente como herramienta para determinar los grados de certeza y confianza que toman las hipótesis y conclusiones. Es importante decir que la estadística es una ayuda para la ciencia pero no es una verdad absoluta (Sarmiento, 2000; Sanjuan, 2006).

Algunas definiciones básicas según Sarmiento (2000), Sanjuan (2006) y Ramírez (2007):

¹ Biólogo Marino. Universidad Jorge Tadeo Lozano. Colombia. EMAIL y MSN biologiamarina@colombia.com

- **Variable:** Tipo de dato, característica o atributo que se toma, mide o se categoriza y que es susceptible de cambiar. (P.ej. densidad de cangrejos en una playa, temperatura, pH, etc.).
- **Observación:** es un dato, una medición de una variable. P.ej. Longitud Total (LT) = 25 cm, Gusanos encontrados = 23, etc.
- **Unidad de muestreo:** es el ejemplar o unidad en que se hace la observación. P.ej.: Pez No. 2., Pez macho No. 45, Gusanos del cultivo 1., etc.
- **Población estadística:** conjunto de unidades de muestreo que se observan o miden. P.ej.: Peces machos, peces hembra, etc.
- **Población biológica:** conjunto de unidades de muestreo totales. Organismos de la especie X. (P.ej.: *Oreochromis niloticus*, *Prochilodus magdalenae*, *Caquetaia kraussii*, *Lutjanus synagris*, *Scomberomorus brasiliensis*, etc.)
- **Muestra:** parte de la población estadística de donde provienen las observaciones. P.ej.: Machos capturados, gusanos contados, etc.
- **Muestreo:** proceso de sacar una muestra.
- **Estadísticos:** valor calculado que describe la variable en la muestra, sirve para analizar e interpretar los datos, de una manera objetiva para llegar a conclusiones veraces de los datos. P.ej.: Promedio de LT, Número de gusanos promedio por cultivo, etc.
- **Potencia:** capacidad de una prueba para detectar una diferencia estadísticamente significativamente.

1.1. Tipos de datos

Esto determina las operaciones y pruebas estadísticas que se pueden aplicar, según Ramírez (2007):

- **Cualitativos:** expresan atributos o cualidades no medibles, pueden ser:
 - **Nominales:** la variable es clasificada por una cualidad propia (atributo) sin una secuencia lógica. (P.ej. sexo, especie, estaciones del año, etc.).

- **Ordinales:** la variable que tiene un orden intrínseco (secuencia lógica). (P.ej. escaso a abundante, pequeño a grande, corto a largo, claro a oscuro, etc.).
- **Cuantitativos:** son datos numéricos que se pueden medir, pueden también ser datos de intervalos. (P.ej. peso (kg), longitud, volumen, velocidades, número de individuos, pH, etc.).

1.2. Tipos de variables

VARIABLES SEGÚN SANJUAN (2006) Y RAMÍREZ (2007):

- **Continuas:** hay cualquier valor concebible entre cada par de datos. (P.ej.: Longitud de un pez, temperatura, pH, etc.).
- **Discreta:** No se presentan valores intermedios entre cada par de datos. (P.ej.: Número de individuos, cantidad de escamas, etc.).
- **Variable dependiente:** es aquella cuyos valores dependen de los que asuma otra variable o factor (variable independiente). (P. ej. Longitud de un renacuajo, número de gusanos encontrados, densidad de *Sphyraena guachancho* dependiente de la densidad de *Opisthonema olginum*), etc.).
- **Factor (variable independiente):** es aquella que, dentro de la relación establecida no depende de ninguna otra. Son las condiciones manipuladas por el investigador a fin de producir ciertos efectos. (P.ej.: el tiempo, pH, O₂, salinidad, profundidad, sitio de muestreo, etc.).

1.3. Tipos de pruebas estadísticas univariadas

TIPOS DE PRUEBAS ESTADÍSTICAS SEGÚN SARMIENTO (2000), SANJUAN (2006) Y RAMÍREZ (2007):

- **Pruebas paramétricas:** tratan exclusivamente con datos numéricos y se basan en la distribución normal.
- **Pruebas no paramétricas:** tratan con datos cualitativos o con numéricos convertidos a ordinales. Son menos exigentes con el tamaño de la muestra, pero son menos potentes. Son usadas cuando los datos no cumplen el tamaño de la muestra, igualdad de varianzas, tipo de distribución, etc.

2. ESTADÍSTICA DESCRIPTIVA

Medidas de tendencia central: estadísticos o parámetros que se sitúan hacia el centro del conjunto de datos ordenados por magnitud (Ramírez, 2007), algunas de estas medidas son:

- **Promedio (media aritmética):** es una medida única, que es muy buena para un análisis descriptivo, pero esta se ve afectada por los valores extremos (Sokal y Rolhlf, 1980; Ramírez, 2007)).
- **Mediana:** punto que divide a la muestra en dos partes iguales (valor de la variable) en una serie ordenada, es también única y no es afectada por los valores extremos (Sokal y Rolhlf, 1980). Teniendo los valores ordenados de mayor a menor o viceversa, el valor medio (en caso de una cantidad impar de muestras n) y el promedio de los dos valores medios (en el caso de una cantidad par de muestras n) es la mediana (Ramírez, 2007).
- **La moda:** valor más frecuente de la distribución de frecuencias o donde la gráfica de esta muestra el pico más alto. Si no se repite ningún valor, no hay moda (Ramírez, 2007).

Gráficas

Las figuras en la estadística descriptiva son un apoyo para interpretar resultados cuantitativos, los tipos de gráficas están relacionados directamente con los tipos de datos, por lo cual no se pueden usar a la ligera.

- **Líneas:** Se utiliza para datos cuantitativos y de variables continuas.
- **Barras:** Para datos cuantitativos y de variables discretas.
- **Pasteles:** Para datos cuantitativos y mostrando porcentajes de variables discretas.

2.1. Medidas de dispersión y variabilidad

Son estimadores de la dispersión de una variable aleatoria de su media. Una valor grande indica que los puntos están lejos de la media, y un valor pequeño indica que los datos están agrupados cerca de la media.

- **Desviación estándar:** es el promedio o variación esperada con respecto a la media en una población. Cuando las poblaciones difieren apreciablemente en sus medias, la comparación de las desviaciones típicas serían bastantes arriesgadas (P. ej. la desviación de las longitudes de la cola de elefantes es obviamente mucho mayor que la longitud de la cola de un solo ratón).
- **Varianza:** Representa la media de las desviaciones. Puesto que están relacionadas con la desviación estándar, cuando las medias de las poblaciones difieren de forma apreciable, no es recomendable usar esta medida de dispersión.
- **Coefficiente de variación:** expresa una variabilidad relativa que compara el grado de variación en poblaciones que tienen diferentes medias, es la desviación estándar expresada como un porcentaje de la media (Sokal y Rohlf, 1980).

3. MÉTODOS UNIVARIADOS

3.1. Distribución normal

Es una distribución de probabilidades que es muy usada en estadística principalmente porque es simétrica y con forma de campana, lo que favorece su aplicación como modelo a gran número de variables (Zar, 1999).

La importancia de la distribución normal, se debe a que muchas de las variables ligadas a la biología siguen el modelo de la normal, como lo son: los caracteres morfológicos de individuos, caracteres fisiológicos, caracteres de comportamiento, entre otros (Zar, 1999) .

3.1.1. Planteamiento de hipótesis

El objetivo principal del análisis bioestadístico es deducir inferencias acerca de la población o ensayo, donde se empieza con establecer una afirmación sobre una “no diferencia” en la hipótesis nula ($H_0: \mu = 0$) y luego se establece la hipótesis alterna, la cual se asume como cierta si se rechaza la hipótesis nula ($H_a: \mu \neq 0$) (Zar, 1999).

NOTA: Debido a que este documento está basado en la utilización del software StatGraphics, para la comprobación de hipótesis en dicho programa se deben invertir las hipótesis para introducirlas en el software, es decir, siempre la igualdad debe ir en la H_a y la contraria en la H_0 .

Según Zar (1999) el criterio para aceptar o rechazar una H_0 es una probabilidad del 5% denominada **nivel de significancia (α)**. El valor del test estadístico correspondiente a α se denomina valor crítico. El α indica cual es la probabilidad de cometer un error tipo I (rechazamos H_0 , siendo verdadera) o la probabilidad de cometer un error tipo II (aceptamos H_0 , siendo falsa) es representada por β .

La naturaleza de la hipótesis alterna determina si la prueba es de una o dos colas, que nos dice la región de rechazo que se concentra a un lado de la curva (Figura 1) o si esta región se divide en los dos lados (Figura 1).

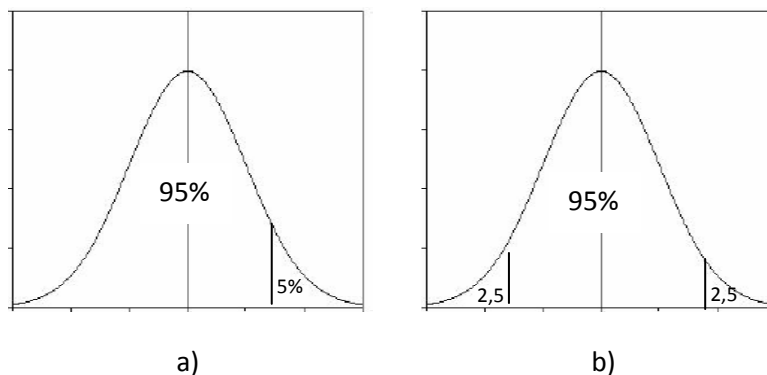


Figura 1. Curva de la distribución normal a) con región de rechazo de una cola y significancia del 95% ($\alpha=0,05$) b) con una región de rechazo de dos colas y significancia del 95% ($\alpha=0,05$). Modificado de Sanjuan (2006).

Dos colas: al investigador NO le interesa si en el muestreo hay una diferencia en una “dirección específica” entre el parámetro y un valor establecido, o entre las medias de dos o más poblaciones. $H_0: \mu = 0$; $H_a: \mu \neq 0$.

Una cola: al investigador le interesa una diferencia en una “dirección específica”. $H_0: \mu \leq 0$; $H_a: \mu > 0$ ó $H_0: \mu \geq 0$; $H_a: \mu < 0$.

3.1.2. Tipos de pruebas estadísticas

- Pruebas paramétricas según Zar (1999):
 - Solo para datos cuantitativos.
 - El tamaño de la muestra (n) igual o mayor a 10 ($n \geq 10$).
 - Se debe ajustar a la distribución normal.
 - Tiene que haber homogeneidad de varianzas de las muestras.
 - El muestreo debe ser aleatorio.
 - Ejemplos: t-Student, t pareada y Análisis de varianza ANOVA.
- Pruebas no paramétricas según Zar (1999):
 - Para datos cuantitativos y cualitativos ordinales.
 - Tamaño de la muestra (n) igual o mayor a 5 ($n \geq 5$).
 - Los datos se pueden ajustar a cualquier distribución (normal, binomial, Poisson, etc.).
 - Con o sin homogeneidad de varianzas.
 - Muestreo debe ser aleatorio.
 - Ejemplos: Mann-Whitney, test de Wilcoxon y de Kruskal-Wallis.

3.1.3. Propiedades de la distribución normal

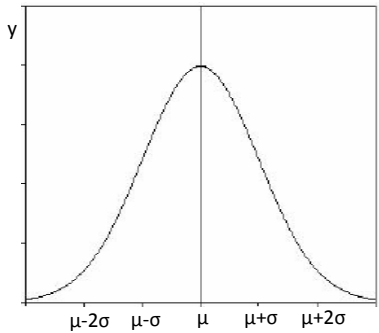


Figura 2. Curva de una distribución normal típica. Modificado de Sanjuan (2006).

Varias distribuciones de diversos datos biológicos tienden a tomar forma de campana típica de normalidad (Figura 2). Los datos tienen una preponderancia a ubicarse alrededor de la media, disminuyendo progresivamente hacia los extremos de los rangos de valores. Es importante notar que no todas las distribuciones en forma de campana se dice que son normales. La altura (Y_i) es la variable dependiente y la independiente es X_i (Zar, 1999; Sanjuan, 2006).

Con la distribución normal se pueden hacer predicciones y tests de hipótesis basados en la suposición de la normalidad y así se puede confirmar o rechazar hipótesis fundamentales

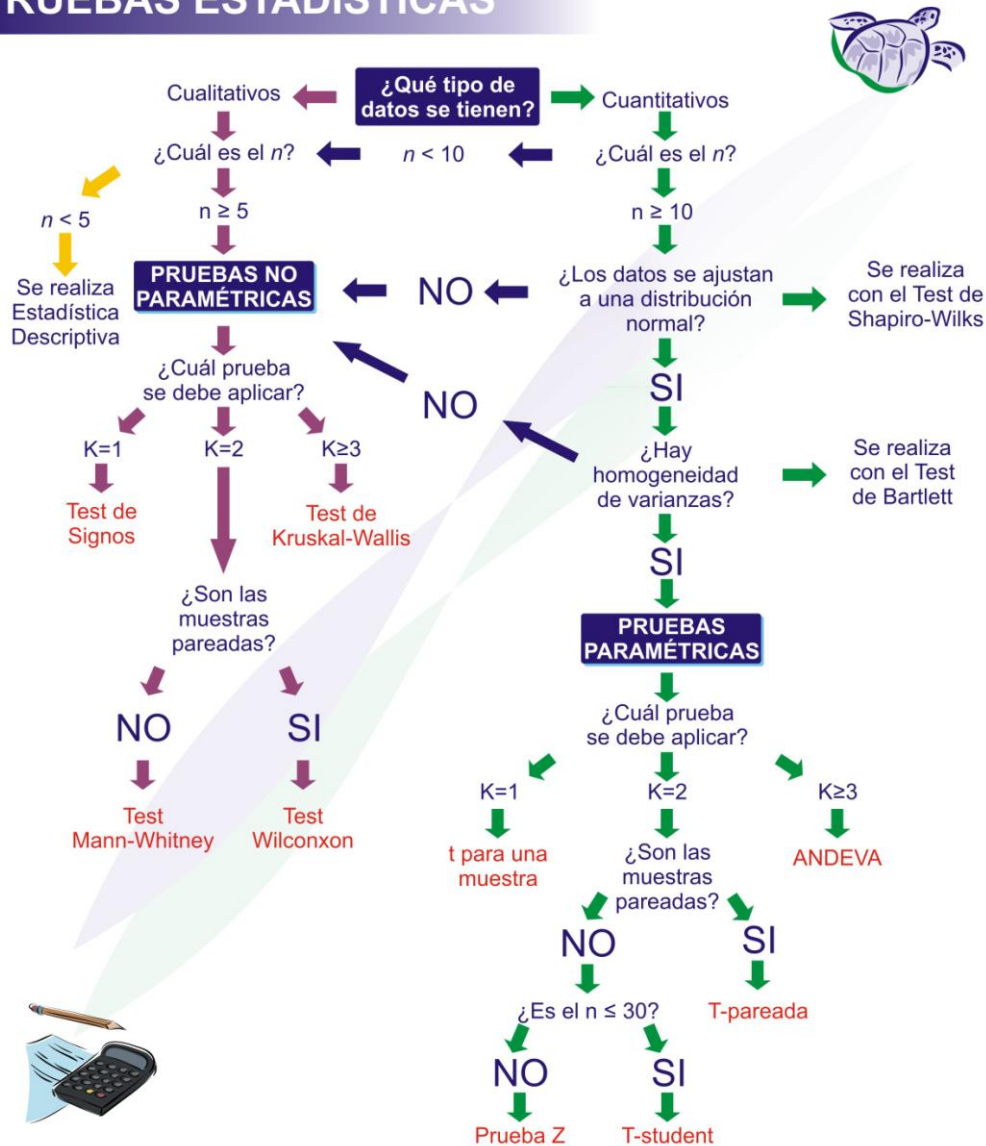
(Zar, 1999; Sanjuan, 2006).

Sin embargo, la distribución se puede separar y generar desviaciones como lo son la asimetría (una cola está más estirada de la otra llamada también sesgo) y la kurtosis (aplastamiento o estrechamiento de la curva) (Zar, 1999; Sanjuan, 2006).

4. EJECUCIÓN DE LAS PRUEBAS ESTADÍSTICAS UNIVARIADAS

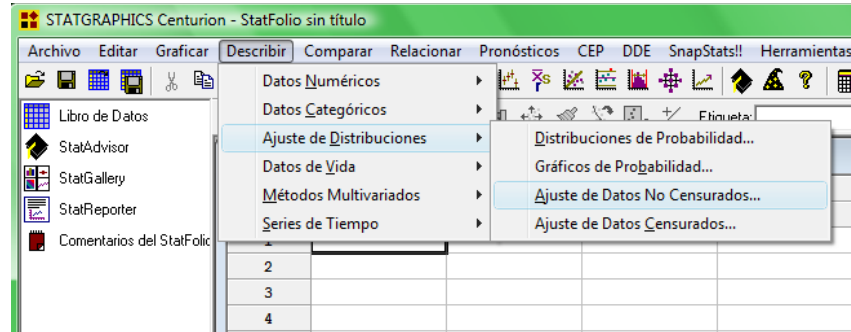
ECOLOGÍA MARINA II - CUADRO GUÍA

PRUEBAS ESTADÍSTICAS

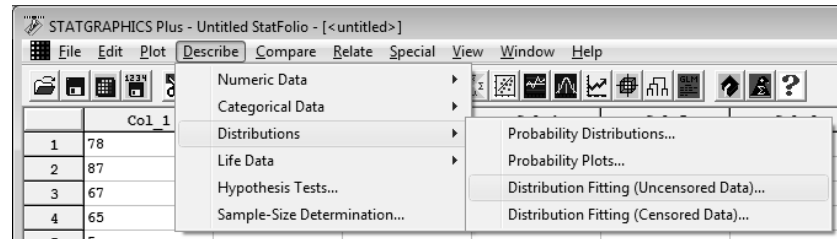


ANDRÉS PULIDO HERNÁNDEZ
 biologiamarina@colombia.com

En el caso de que los datos sean cuantitativos, y el $n \geq 10$, se procede a ver si los datos se ajustan a la distribución normal, para ello se puede usar el test de Shapiro-Wilks (1965) basado en el cálculo del estadístico W (Figura 3) el cual según Sanjuan (2006) tiene una potencia excelente pero se ve afectado cuando hay muchos datos idénticos.



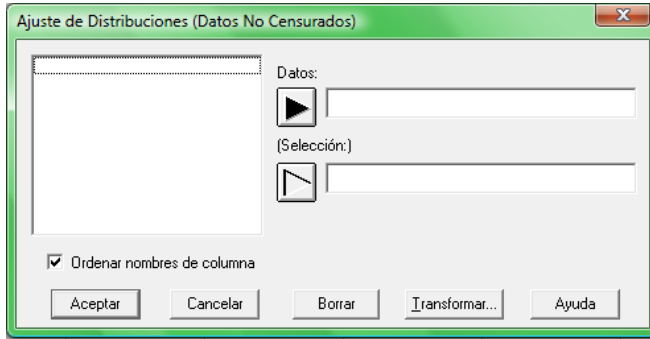
a)



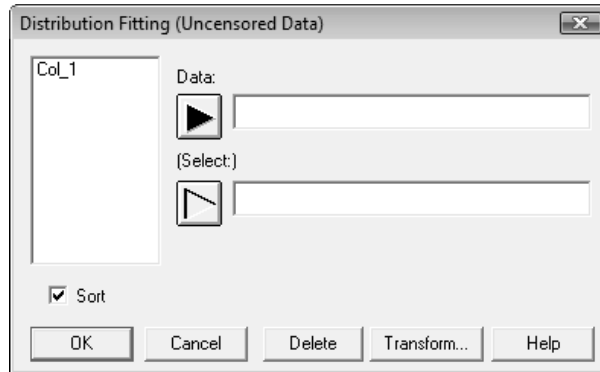
b)

Figura 3. Pasos a seguir en el software StatGraphics para evaluar el ajuste a la normalidad de un conjunto de datos a) Versión XV Centurion b) versión 4.0 y 5.0.

Pasos a seguir: Describir (Describe), Ajuste de distribuciones (Distributions), Ajuste de datos No censurados (Distribution Fitting (Uncensored Data))...





a)

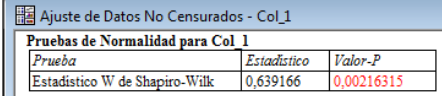


b)

Figura 4. Selección del conjunto de datos en la ventana de la izquierda, donde cada variable. También en esta misma ventana se pueden hacer transformaciones a los datos en caso de necesitarlas con el botón “Transformar” previamente indicando la variable a trabajar a) Versión XV Centurion b) Versión 4.0 y 5.0.

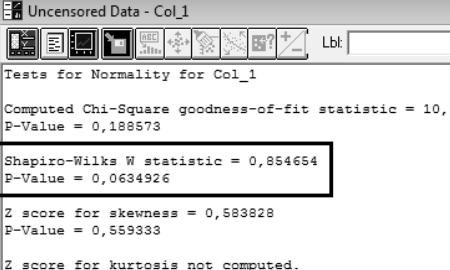
Se escoge la columna (variable de datos) que se va a evaluar. NOTA: Para cada muestreo se debe realizar el mismo procedimiento y se da click en Datos (Data), luego en OK (Figura 4).

Luego se seleccionan las opciones tabulares en el icono amarillo de arriba a la izquierda  (SG 4,0 o 5,0) o  (SG XV) y se habilita la opción **Pruebas de Normalidad (Test for Normality)** y se deshabilita opcionalmente **Resumen de Análisis (Analysis Summary)**. Y luego se da doble click sobre la ventana que aparece para maximizarla.



Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0,639166	0,00216315

a)



```

Tests for Normality for Col_1

Computed Chi-Square goodness-of-fit statistic = 10,0
P-Value = 0,188573

Shapiro-Wilks W statistic = 0,854654
P-Value = 0,0634926

Z score for skewness = 0,583828
P-Value = 0,559333

Z score for kurtosis not computed.
  
```

b)

Figura 5. Resultado del test de Shapiro-Wilks a) Versión XV Centurion b) versión 4.0 y 5.0.

En la ventana aparecen una serie de resultados, donde el segundo que se muestra es la **Prueba de Shapiro-Wilks** (Test) y nos indica el valor P (P-Value) (Figura 5) que debe ser menor o igual a 0,05

para ajustarse a una distribución normal, si es mayor a 0,05 no se ajustará a la normalidad, pero queda la opción de hacer una serie de transformaciones a los datos según la naturaleza de estos.

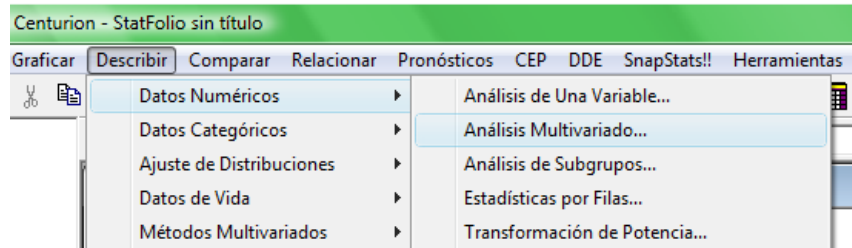
Transformación de datos:

Los datos se pueden transformar según el comportamiento que tengan por el cual no se hayan ajustado a la distribución normal, donde se encuentran 4 casos según Zar (1999) y Sanjuan (2006):

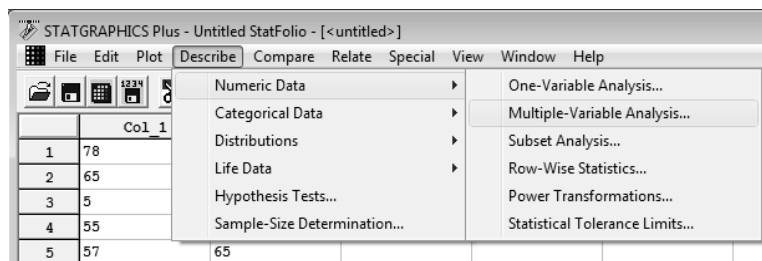
- Si hay sesgo (desplazamiento de la curva normal a la derecha o izquierda) entonces se debe transformar con *logaritmo* que estira la cola de la distribución.
- Si hay leptocurtosis (ubicación de la mayoría de los datos cercanos a la media), algunos datos muy bajos o ceros se debe transformar con *raíz cuadrada*.
- Si hay platicurtosis (dispersión de los datos a través de todo el eje X) los datos no se pueden transformar y se debe hacer *pruebas no paramétricas*.
- Si hay ceros dentro de los datos del muestreo, se debe transformar con *log (x+1)*.
- Cuando son porcentajes se transforma con *arcsen*.

Para evaluar si hay sesgo o curtosis (platicurtosis o leptocurtosis) se deben seguir los siguientes pasos QUE SON LOS MISMOS PARA LA ESTADISTICA DESCRIPTIVA COMO MEDIA, MODA, MEDIANA, ETC.:

Pasos a seguir: Describir (Describe) > Datos numéricos (Numeric Data) > Análisis multivariado (Multiple-Variable Analysis)... (Figura 6)





a)



b)

Figura 6. Pasos a seguir en el software StatGraphisc para estadística descriptiva a) Versión XV Centurion b) versión 4.0 y 5.0.

Luego aparece una ventana (Figura 7) con las variables a considerar, se seleccionan todas y se da click en la flecha de Data para pasarlas al cuadro de la derecha. Luego se da click en OK (recomendado dejar habilitado el cuadro de Sort). Luego cuando aparece la nueva ventana dar click en el botón amarillo de la izquierda en la barra de herramientas Opciones Tabulares  (SG 4,0 o 5,0) o  (SG XV) (Tabular options) y habilitar la casilla de “Resumen estadístico (Summary Statistics)”. Si la cantidad de datos entre variables son diferentes entonces dar click derecho sobre la ventana blanca y dar click en “Opciones de Análisis (Analysis Options...)” y seleccionar “Todos los datos (All Data)”. Luego volver a dar click derecho sobre la ventana blanca pero ahora dar click en

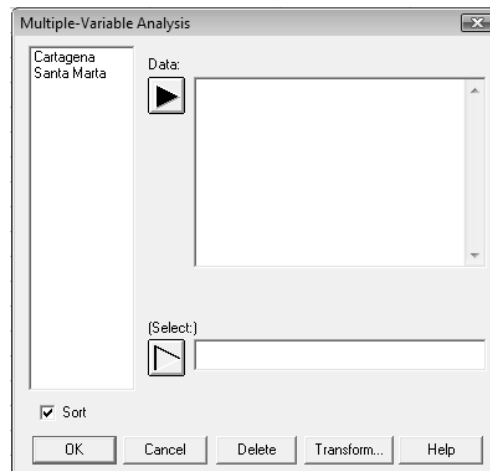
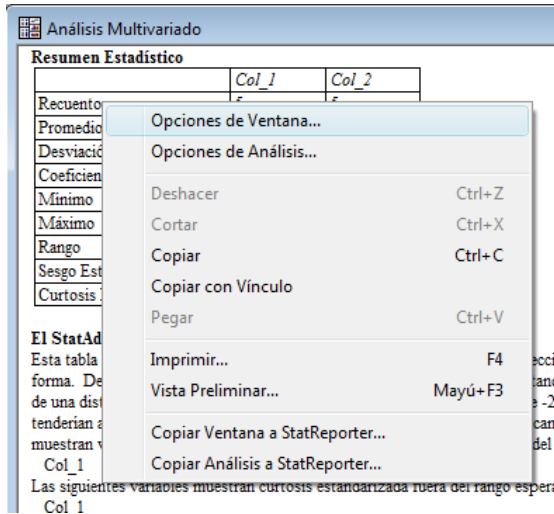


Figura 7. Selección de variables para estadística descriptiva.

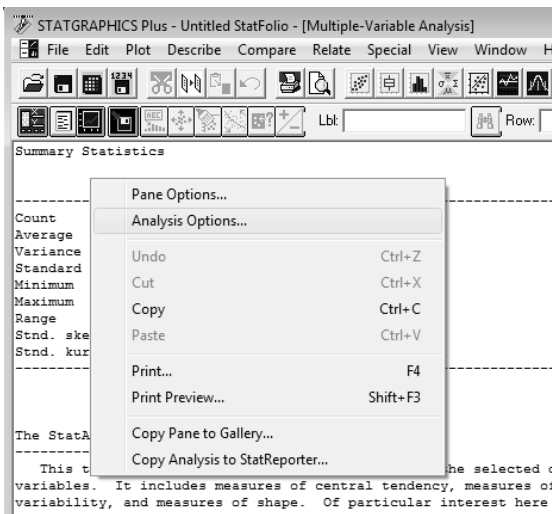
“Opciones de Ventana (Pane Options...)” (Figura 8), sale un cuadro y habilitar la opciones de: media (average), mediana (median), moda (mode), varianza (variance), desviación estándar (std. Deviation), mínimo (min), máxima (max), sesgo (std. Skewness) (sesgo), kurtosis (std. Kurtosis) y coeficiente de variación (Coeff. Of Var.) como se observa en la Figura 9.



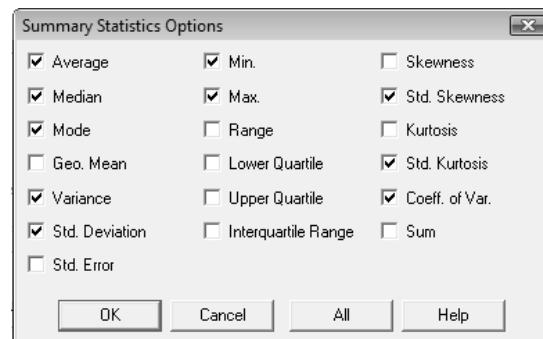
a)



a)



b)

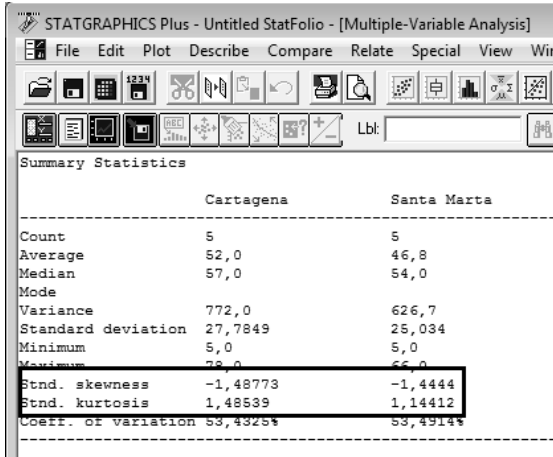


b)

Figura 8. Selección de todos los datos de las variables) Versión XV Centurion b) versión 4.0 y 5.0.

Figura 9. Selección de medidas de tendencia central a) Versión XV Centurion b) versión 4.0 y 5.0.

Con esta tabla, se observan los valores de sesgo y curtosis, además de las otras medidas de tendencia central. En el caso del sesgo, para que no haya debe estar entre -2 y 2, si es mayor a 2 hay un sesgo positivo y si es menor a -2 hay un sesgo negativo. En el caso de la curtosis, si está entre -2 y 2 no hay curtosis, si es mayor a 2 es leptocúrtica y si es menor a -2 es platicúrtica (Figura 10).

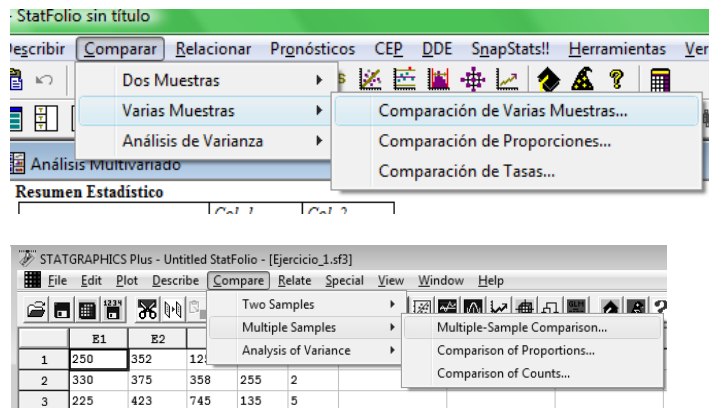


	Cartagena	Santa Marta
Count	5	5
Average	52,0	46,8
Median	57,0	54,0
Mode		
Variance	772,0	626,7
Standard deviation	27,7849	25,034
Minimum	5,0	5,0
Maximum	78,0	66,0
Std. skewness	-1,48773	-1,4444
Std. kurtosis	1,48539	1,44412
Coeff. of Variation	53,4325%	53,4914%

Figura 10. Resultados de algunas medidas de tendencia central, sesgo y kurtosis.


Según Sanjuan (2006) para comparar dos o más muestras con estadística paramétrica es necesario comprobar que las muestras provienen de un conjunto de datos con idénticas varianzas. Hay diferentes métodos para determinar esto, uno de los más usados es la **Prueba de Bartlett** al ser uno de los más potentes pero es muy dispendioso. El resultado del valor P (P-value) debe ser mayor o igual a 0,05 para que haya homogeneidad de varianzas.

Pasos a seguir: Comparar (Compare), Varias Muestras (Multiple Sample), Comparación de Varias Muestras (Multiple-Sample Comparison...) (Figura 11).



Seguidamente aparece una pequeña ventana, donde se debe seleccionar "Multiple columnas de datos (Multiple Data Columns)" y se da click en ACEPTAR (OK) (Figura 12).

Figura 11. Pasos a seguir en el software Stat Graphics para evaluar la homogeneidad de varianzas a) Versión XV Centurion b) versión 4.0 y 5.0.

Luego se seleccionan las columnas a comparar hasta que todas queden sombreadas y se da click en la flecha negra, luego un click en OK. Después se da click en el botón amarillo de arriba a la izquierda “Tabular Options ” y se habilita la opción de “Varianza Check”. Se puede dar un doble click a la nueva ventana que aparece donde estará el resultado del Test de Bartlett (Figura 13).

El resultado de esta Prueba, se mira el valor P (P-Value), si este es mayor a 0,05 hay homogeneidad de varianzas.

Si se cumple todo esto se pueden hacer PRUEBAS PARAMÉTRICAS. Para realizar la prueba que mejor se aplique, (ver diagrama de flujo.pdf) ir al documento “Métodos Univariados.pdf”.

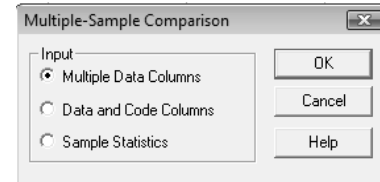
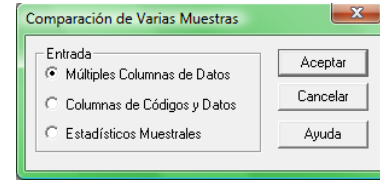


Figura 12. Selección de la comparación de múltiples muestras a) Versión XV Centurion b) versión 4.0 y 5.0.

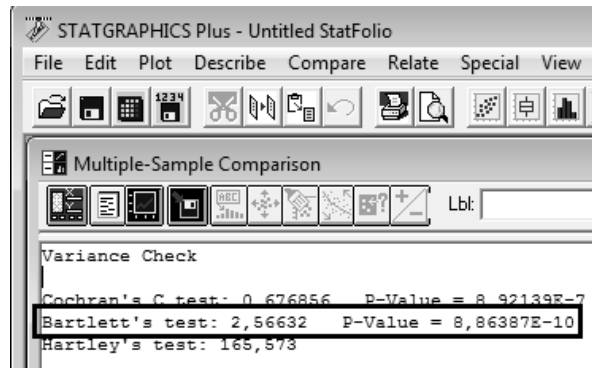


Figura 13. Resultado del Test de Bartlett que se compara con la Tabla χ^2 .

5. MÉTODOS MULTIVARIADOS

Los métodos estadísticos multivariados son herramientas para hacer inferencias de datos cuantitativos o algunas veces cualitativos de comunidades biológicas, utilizando instrumentos como los índices de diversidad (dominio, uniformidad, riqueza, etc.), análisis de clasificación (clusters, MDS, etc). Estos índices surgieron por la necesidad para descifrar el funcionamiento de las comunidades y ecosistemas en expresiones matemáticas que reflejaran una relación entre número de especies y la proporción de sus individuos (Clarke y Gorley, 2001; Ramírez, 2005).

5.1. Índices de Diversidad

5.1.1. Instrucciones del manejo del programa PRIMER v5.0 (Clarke y Gorley, 2001)

Ordenar en una hoja de cálculo (preferiblemente Microsoft Excel versión 2003 o anterior) los datos colocando en la primera columna los organismos (especies, géneros, familias, etc.). En la primera fila los estratos a comparar (estaciones, transectos, parcelas, ecosistemas, etc.). En las celdas de datos, es importante que sean todos numéricos y las que se encuentren en blanco reemplazarlas con cero.

Seguidamente en el PRIMER se abre la matriz a trabajar (**File > Open > Excel Files**), se deja en blanco la selección de "Includes title" y click en OK. Luego, opcionalmente se agrega un título y click en OK (Figura 14).

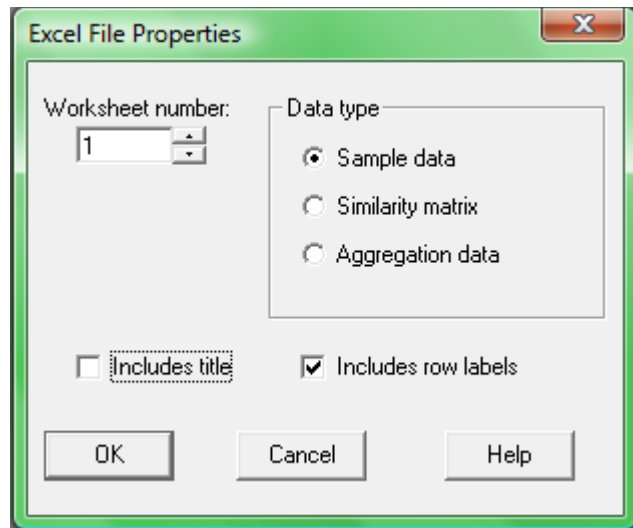
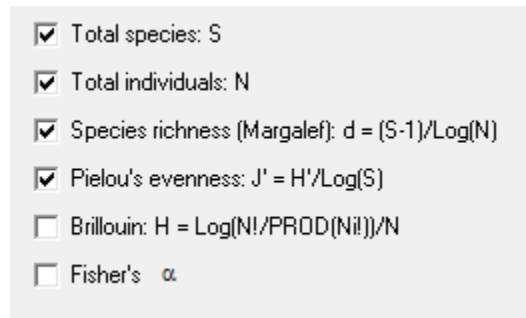


Figura 14. Ventana propiedades del archivo en Excel.

Ya teniendo la matriz, seguidamente se realiza el cálculo de los respectivos índices (Analyse > Diverse...). La ventana siguiente muestra los diferentes índices a calcular en viñetas en la parte superior, en casos básicos se seleccionan los siguientes:

Other | Shannon | Simpson | Hill | Rarefaction | Taxdisc | Phylogenetic |

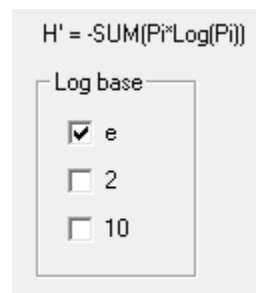
Pestaña Otros: Especies totales (S) indica el número total de la suma de especies encontradas, Individuos totales (N) indica la abundancia total de organismo, Riqueza de Margalef (d), Uniformidad de Pielou (J') muestra un valor en porcentaje que es análogo a la equiparabilidad que determina la diversidad de una muestra (Figura 15).



- Total species: S
- Total individuals: N
- Species richness (Margalef): $d = (S-1)/\text{Log}(N)$
- Pielou's evenness: $J' = H'/\text{Log}(S)$
- Brillouin: $H = \text{Log}(N!/\text{PROD}(N_i!))/N$
- Fisher's α

Figura 15. Ventana índices de la pestaña "otros".

Pestaña Shannon (Diversidad): hay tres logaritmos con bases diferentes para calcular, en la actualidad se tiende a manejar en base e, sin embargo, lo importante es que todos los resultados se calculen con el mismo logaritmo para que sean comparables los datos, y si por ejemplo los resultados de la investigación se van a comparar con otros trabajos, se debe calcular éste índice con la base del logaritmo de los demás realizados (Ramírez, 2005) (Figura 16).



$H' = -\text{SUM}(P_i \cdot \text{Log}(P_i))$

Log base

- e
- 2
- 10

Figura 16. Ventana índices de la pestaña "Shannon".

Pestaña Simpson (Predominio): indica la posible dominancia de una o varias especies dentro de la muestra, tiene cuatro alternativas a seleccionar, sigma se utiliza como predominio de Simpson, 1-sigma es diversidad de Simpson (poco utilizado porque se ve muy marcado dependientemente por las especies abundantes). La sigma prima se utiliza en el caso de comunidades estadísticamente finitas, y en la sigma sencilla es para grupos infinitos (Ramírez, 2005).

En estos casos, se recomienda usar sigma o 1-sigma, no al tiempo ya que ambas muestran el mismo resultado interpretado de forma inversa, es decir, mientras el primero indica “dominio” el segundo indica “uniformidad” (Figura 17).

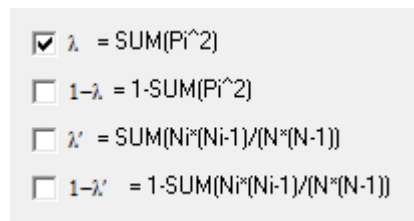


Figura 17. Ventana índices de la pestaña “Simpson”.

Pestaña Hill: se trabajan básicamente los números de Hill 1 y 2, el primero indica la cantidad de especies abundantes y el segundo la cantidad de especies muy abundantes (Figura 18).

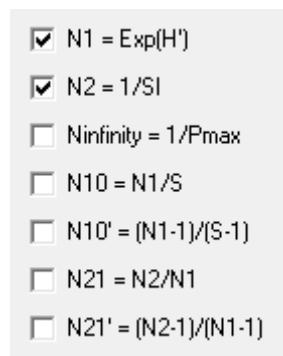


Figura 18. Ventana de índices de la pestaña “Hill”.

Por último en la misma tabla se recomienda dar click en el cuadro de “Results in worksheet” para que se puedan apreciar en una forma ordenada dentro de una matriz, la cual seguidamente se

puede copiar y pegar en una tabla de Microsoft Excel para un posterior tratamiento u organización de datos.

5.1.2. Análisis de matriz de diversidad

Por lo general, cuando se habla de un ecosistema diverso se deben tener en cuenta la composición de especies con una alta riqueza, alta abundancia, un bajo predominio, una alta uniformidad. Estas son las condiciones ideales, sin embargo, suelen fluctuar variando algunas o todas.

Para comparar varias muestras es necesario observar y analizar de forma conjunta varios índices para determinar balances entre unos y otros, y así tener conclusiones más certeras basadas en estas herramientas estadísticas.

En primera instancia, se explica a continuación los valores entre los cuales fluctúa cada índice y su interpretación individual:

- La **Riqueza (S)** determina la cantidad de especies dentro de la muestra y la **Abundancia relativa (N)** muestra el valor de la cantidad de organismos de una especie determinada. La **Riqueza de Margalef (d)** es un índice que se utiliza para relacionar las dos anteriores en un solo valor, oscilando entre 0 e infinito, donde 0 es una baja riqueza (Ramírez, 2005).
- El índice de **Shannon (H')** asume comunidades infinitamente grandes que no se pueden estudiar en su totalidad por infraestructura, costos, tiempo, personal, por tal razón su valor debe estimarse a partir de una muestra; así el índice va desde 0 hasta infinito, éste valor hace una inferencia primaria sobre la diversidad de la muestra, la cual se apoya en gran forma con el de **Pielou (J')** el cual determina un resultado semejante, pero que es comparado con la diversidad máxima y así arroja un valor entre 0 y 1 (o porcentaje) que es comparable y más potente que el índice de Shannon solo, en Pielou 1 es la máxima diversidad y 0 la mínima.
- El intervalo del índice de **Simpson (Lambda)** es conocido como una medida de concentración y se refiere a la probabilidad de extraer dos individuos de una misma especie en una muestra, también sirve como una medida de diversidad, pero como es muy marcado como dependiente de las especies más abundantes tiene más fuerza como un índice de dominancia (Ramírez, 2005); va de 0 a 1, donde 0 es bajo predominio, lo que sugiere una uniformidad alta entre las abundancias de las especies en la muestra, y cuando se alcanza o está cerca de 1, determina un predominio alto por una o varias especies dentro del total.

5.2. Análisis de clasificación

5.2.1. Cluster

Es una representación gráfica de dos dimensiones a través de dendogramas (Figura 19), expresando el grado de semejanza entre dos o más conjuntos multivariados. Se aplican en matrices tipo Q (especies en las filas y estaciones en las columnas, donde estas últimas son las entidades a comparar y las especies los atributos) o tipo R (especies en las columnas y estaciones en las filas, donde se realizan asociaciones entre especies que son las entidades y las estaciones los atributos), entre las interacciones entre filas y columnas se encuentran los coeficientes de afinidad; al tener el dendograma listo, luego se procede a establecer el **Índice de Afinidad** para aceptar una estación dentro de un grupo establecido en la gráfica, la literatura no da respuesta a esto, por lo cual no hay un valor establecido para tal situación. No obstante, dado que una afinidad da 0,5 indica más semejanza que diferencia entre los grupos que se comparan (Figura 20), este valor constituye un umbral a tener en cuenta como punto de partida, aunque es algo flexible alrededor de este, por ejemplo podría fijarse en 0,6 o 0,7 (Ramírez, 2005).

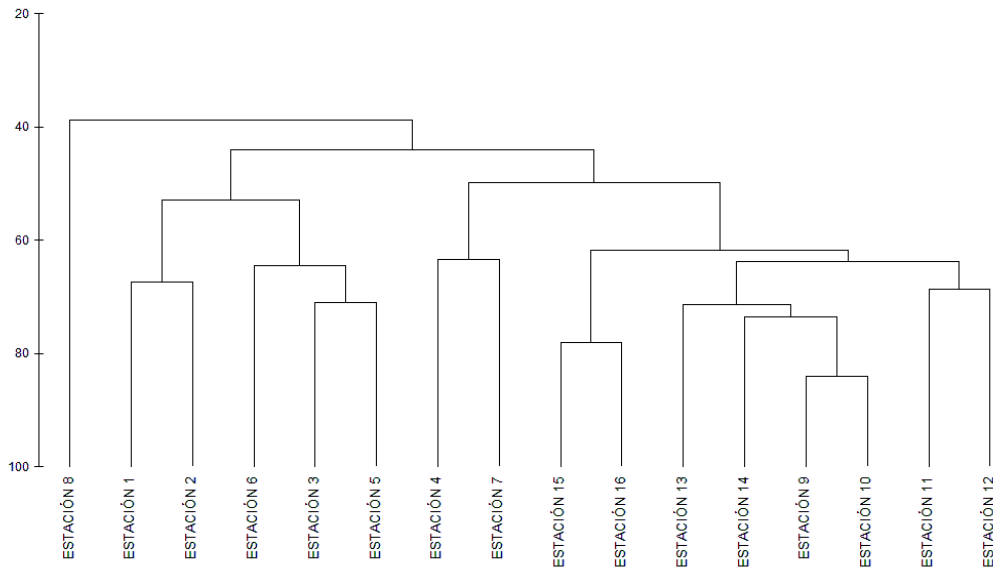


Figura 19. Dendograma sin Índice de Afinidad.

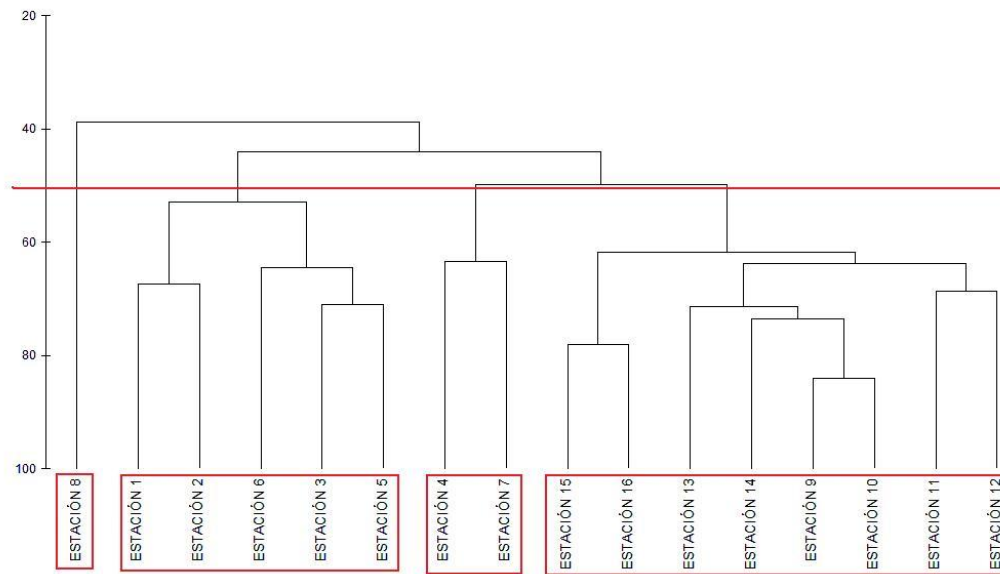


Figura 20. Dendrograma con Índice de Afinidad de 50% = 0,5.

5.2.2. MDS

Adicionalmente, otra representación gráfica de dos dimensiones que expresa cualitativamente el grado de semejanza entre dos o más conjuntos multivariados es el MDS, el cual muestra cercanías entre entidades a comparar, en matrices tipo Q estaciones y en las de tipo R afinidad de especies. Para que este diagrama sea representativo el valor de estres debe ser menor a 0,01, en caso contrario no es muy confiable.



Figura 21. Representación gráfica del MDS.

6. BIBLIOGRAFÍA RECOMENDADA

- Castillo, I. 2006. Estadística descriptiva y cálculo de probabilidades. Pearson Educación. Madrid. 425 p. 519.2 C278
- Clarke, K. y R. Gorley. 2001. PRIMER v5. User Manual/Tutorial. PRIMER-E Ltd. United Kingdom. 91 p.
- Cristófoli, M. 2007. Manual de estadística con Microsoft Excel. Omnicrom System. Buenos Aires. 240 p. 519.5 C865
- Horra, J. 1995. Estadística aplicada. Ediciones Días de Santos. Madrid. 179 p. 519.5 H816
- Martínez, C. 2005. Estadística y muestreo. Ecoe Ediciones. 12ª edición. Bogotá. 998 p. 519.5 M17EST 2005
- Milton, J. 2001. Estadística para biología y ciencias de la salud. McGraw Hill Interamericana. 3ra. Edición. Madrid. 592 p. 574.0182 M662
- Ramírez, A. 2005. Ecología aplicada. Diseño y análisis estadístico. Fundación Universidad Jorge Tadeo Lozano. Bogotá. 325 p. 574.5 R173 2005
- Ramírez, A. 2007. Ecología. Introducción a la aplicación matemática. Pontificia Universidad Javeriana. Fundación Cultural Javeriana. Bogotá. 224 p. 574.5 R173E
- Sokal R. y J. Rohlf. 1980. Introducción a la Bioestadística. Editorial Reverté S.A. Barcelona. 362 p. 574.0182 S683
- Zar J. 1999. Biostatistical Analysis. Prentice Hall. 4ª edición. New Jersey. 663 p. 570.1 Z36

7. BIBLIOGRAFÍA CITADA

- Clarke, K. y R. Gorley. 2001. PRIMER v5. User Manual/Tutorial. PRIMER-E Ltd. United Kingdom. 91 p.
- Ramírez, A. 2005. Ecología aplicada. Diseño y análisis estadístico. Fundación Universidad Jorge Tadeo Lozano. Bogotá. 325 p.
- Ramírez, A. 2007. Ecología. Introducción a la aplicación matemática. Pontificia Universidad Javeriana. Fundación Cultural Javeriana. Bogotá. 224 p.
- Sanjuan A. 2006. Estadística aplicada a la ecología. Presentaciones digitales del curso de Ecología II. Facultad de Biología Marina. Universidad Jorge Tadeo Lozano. Santa Marta.
- Sarmiento, F. 2000. Diccionario de ecología. Paisajes, conservación y desarrollo sustentable para Latinoamérica. Quito. 362 p.
- Sokal R. y J. Rohlf. 1980. Introducción a la Bioestadística. Editorial Reverté S.A. Barcelona. 362 p.
- Zar J. 1999. Biostatistical Analysis. Prentice Hall. 4ª edición. New Jersey. 663 p.