**Ejercicio Semanas 12-13 (Normalidad) – Estadística 2019-II**

**Autores:** Maria Alejandra Navarro – 2161578; Pablo Emilio Mojica Pradilla – 2160008

Lo primero que se hizo fue una observación y exploración básica de los datos “Morphological\_data.csv”, escogiendo aquellas variables relacionadas con México y trabajando con ellas según las instrucciones y ejemplos que se mostraban en la práctica de laboratorio.

Como segundo paso, se realizó una observación gráfica de las variables “BL”, “WB”, “DB” y “WLLP”.

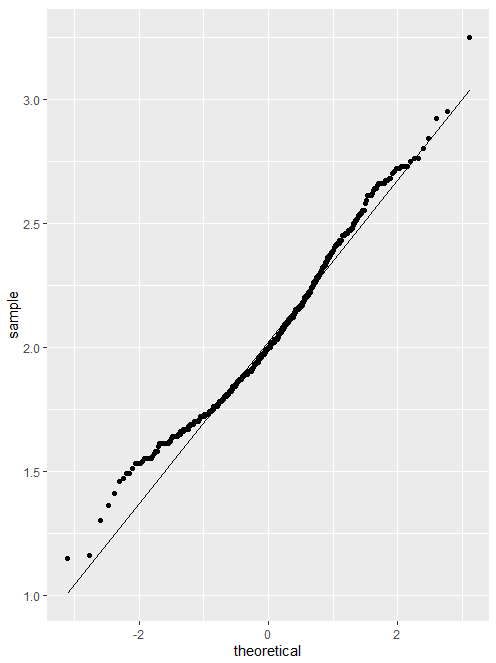
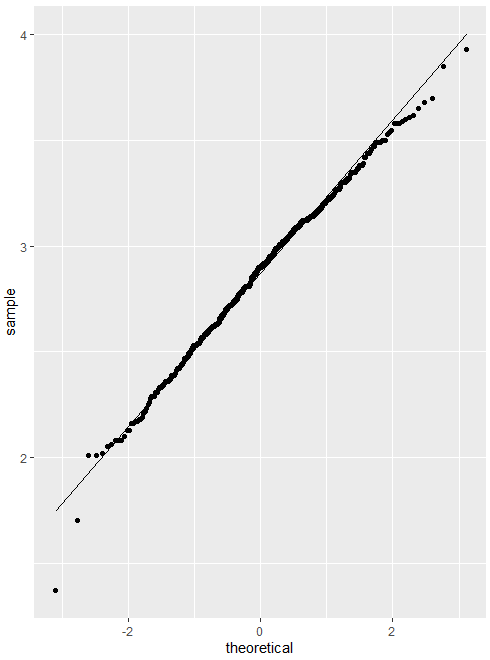
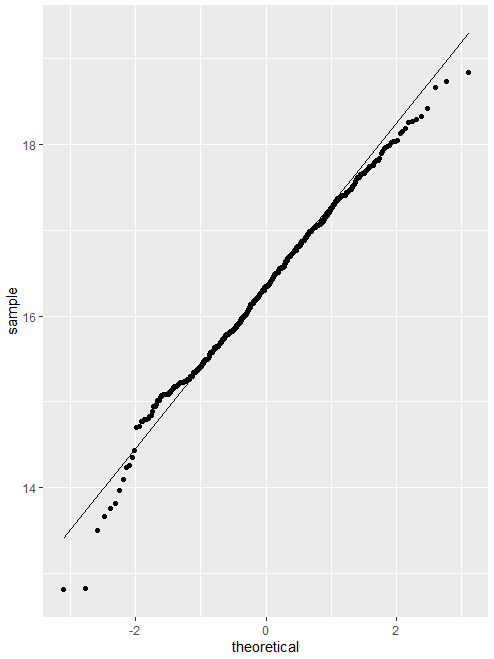
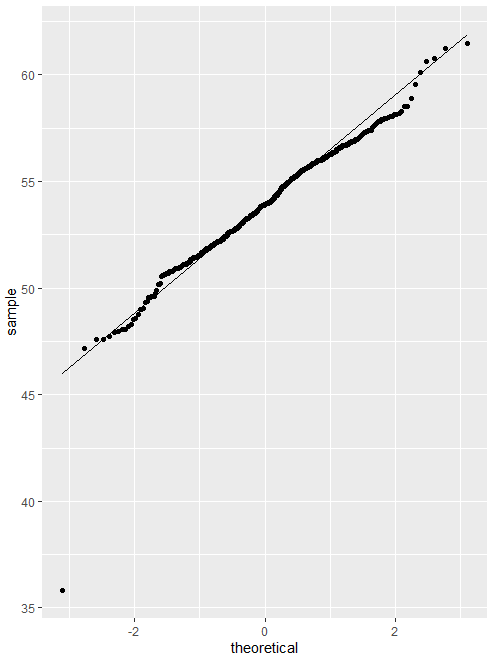
 

Figura 1. Diagramas de dispersión de las variables “BL”, “WB”, “DB” Y “WLLP”, respectivamente.

Posteriormente se aplicó un test de normalidad Lillie test (dado que habían más de 50 datos) a cada conjunto de datos con el fin de determinar si seguían un comportamiento normal o no. Según los resultados obtenidos, “BL” y “WLLP” son normales (p-valor de 0.031 y 0.113, respectivamente), mientras que “WB” y “DB” no son normales (p-valor de 0.041 y 0.00002, respectivamente). En cuanto a la curtosis, las 4 variables son leptocúrticas.

Luego de realizar el Lillie test, se realizó un test de homogeneidad de varianzas (homocedasticidad) a las variables; se utilizó el Levene test. “BL” (p-valor de 0.0056) es la única variable que no presenta homogeneidad de varianza, pues “WB”, “DB” y “WLLP” sí presentan (p-valor de 0.185, 0.301 y 0.284, respectivamente).

En base a los análisis de normalidad, se les realizó una transformación a las variables “WB” y “DB” con el fin de ver si se podían transformar a distribuciones normales. Se realizó transformación de Tukey y solamente la variable “WB”, luego de ser transformada, pasó a ser normal (p-valor de 0.853); la variable “DB”, aún siendo transformada, siguió con un comportamiento no normal (p-valor de 0.0063).

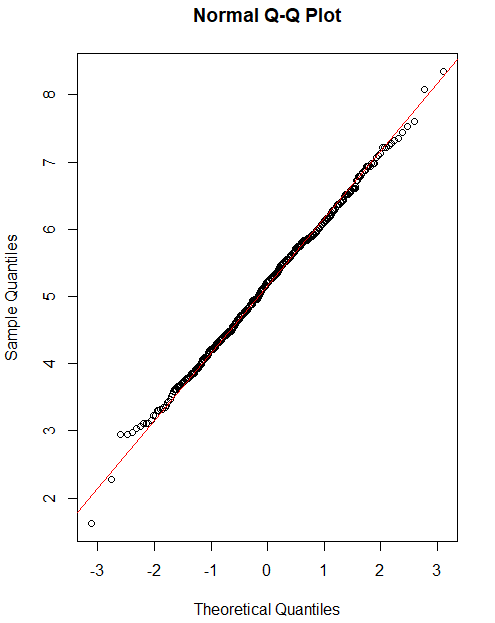
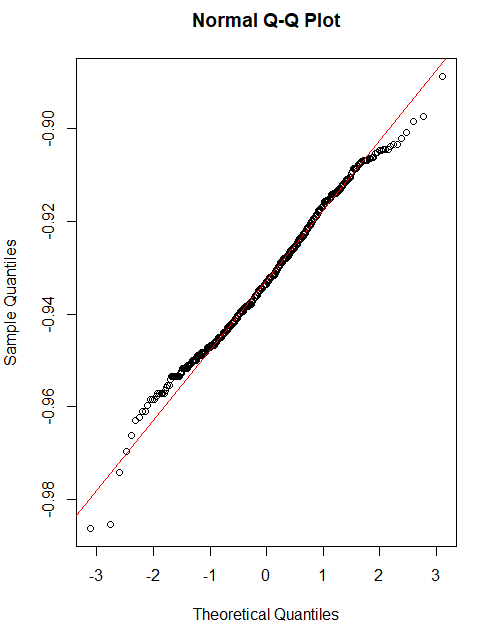
 

Figura 2. Diagramas de dispersión de las variables “WB”, “DB”, respectivamente, luego de la transformación.

Por último, conociendo el comportamiento de normalidad para la distribución de las variables trabajadas, se realizaron algunos tests paramétricos (para las normales) y no paramétricos (para las no normales). Se utilizaron los datos de la variable “BL” para México y Estados Unidos (distribuciones normales), realizando un t test de “dos lados”, “mayor” y “menor”; el t test de “dos lados” demostró que las medias son diferentes (p-valor de 0.00151), el t test “mayor” demostró que la media de x es menor (p-valor de 0.999), y, por último, el t test “menor” demostró que y es mayor (p-valor de 0.00075). En cuanto al test de comparación de varianzas, éste demostró que las varianzas son diferentes (p-valor de 4.35). Por otro lado, para los tests no paramétricos se utilizaron los datos de la variable “WB” para México y Estados Unidos (distribuciones no normales), realizando un Wilcox test de “dos lados”, “mayor” y “menor”; el Wilcox test de “dos lados” demostró que no hay diferencias entre los datos (p-valor de 0.066), el Wilcox test “mayor” demostró que la media de x es mayor (p-valor de 0.033), y, por último, el Wilcox test “menor” demostró que la media de y es menor (p-valor de 0.967).

Según los tests y análisis realizados, puede decirse que las variables estudiadas para México en la matriz de datos morfológicos son altamente variables. Se tienen tanto distribuciones normales como no normales y dentro de las no normales hay algunas que pueden transformarse en normales y otras que no (tienen comportamientos muy alejados de la normalidad). También puede decirse que los tamaños de las muestras sin dudas tienen un efecto importante sobre los resultados obtenidos a lo largo de la práctica, sin embargo, por ejemplo, las variables analizadas tienen homogeneidad de varianza a nivel general (sólo una variable de cuatro analizadas, no tiene). Lo visto durante los temas de normalidad nos permitió entender cómo pueden ser los distintos comportamientos en un mismo grupo de datos y lo que se puede hacer con ellos, desde pruebas de normalidad hasta tests paramétricos o no paramétricos según la naturaleza de la distribución.