**Trabajo Práctico N° 4**

**Ejercicio 1.**

*Construir un test de normalidad multivariada y aplicarlo sobre los datos correspondientes al primer ejercicio de la práctica 2. ¿Se puede concluir que el vector de medias de los datos es distinto del vector nulo? Recordemos del Ejercicio 1 de la Práctica 2 que el archivo “ine.dta” es una base de datos que contiene los gastos promedio, en euros, de los hogares españoles, por grandes rubros y comunidad autónoma, correspondientes a los relevamientos de la encuesta de presupuestos familiares del año 2005 realizada por el Instituto Nacional de Estadísticas de España (INE).*

Test for multivariate normality

Mardia mSkewness = 84.03333 chi2(286) = 302.066 Prob>chi2 = 0.2459

Mardia mKurtosis = 130.2492 chi2(1) = 2.558 Prob>chi2 = 0.1097

Henze-Zirkler = .9871703 chi2(1) = 3.025 Prob>chi2 = 0.0820

Doornik-Hansen chi2(22) = 18.126 Prob>chi2 = 0.6985

Por lo tanto, con un nivel de significancia del 10% (excepto con el test Henze-Zirkler), estos datos no aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

Test that all means are 0

Hotelling T2 = 1894.51

Hotelling F(11,7) = 70.92

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias de los datos es distinto del vector nulo.

**Ejercicio 2.**

*Generar una muestra de 100 observaciones de una población normal multivariada teniendo en cuenta los siguientes vectores de parámetros:*

*= , = .*

**(a)** *Efectuar un test para verificar la normalidad de la base de datos construida.*

Test for multivariate normality

Mardia mSkewness = 8.975172 chi2(120) = 155.098 Prob>chi2 = 0.0171

Mardia mKurtosis = 81.19782 chi2(1) = 0.224 Prob>chi2 = 0.6359

Henze-Zirkler = .9600012 chi2(1) = 0.093 Prob>chi2 = 0.7601

Doornik-Hansen chi2(16) = 16.982 Prob>chi2 = 0.3867

Por lo tanto, con un nivel de significancia del 10% (excepto con el test de Mardia mSkewness), estos datos no aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

**(b)** *¿Existe evidencia empírica suficiente que permita concluir que el vector de medias es, estadísticamente, distinto del vector nulo?*

Test that all means are 0

Hotelling T2 = 2798.76

Hotelling F(8,92) = 325.11

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias no es igual al vector nulo.

**(c)** *Efectuar un test para contrastar la hipótesis acerca de la igualdad de las medias de las variables que componen la matriz de datos.*

Test that all means are the same

Hotelling T2 = 907.75

Hotelling F(7,93) = 121.82

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

**(d)** *Contrastar la esfericidad de la distribución de los datos.*

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 412.74

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

**(e)** *¿Se verifica, estadísticamente, la diagonalidad de la matriz de varianzas y covarianzas en los datos?*

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 21.96

Prob > chi2 = 0.7833

Por lo tanto, con un nivel de significancia del 10%, estos datos no aportan evidencia suficiente para indicar que la matriz de datos no es diagonal.

**Ejercicio 3.**

*Repetir el ejercicio anterior, considerando, en este caso, la siguiente estructura:*

*X= , (, ), (, ), (, ),*

*siendo:*

*= , = , = 40%;*

*= , = , = 50%;*

*= , = , = 10%.*

**(a)** *Comentar lo observado en contraste con el ejercicio anterior.*

Test for multivariate normality

Mardia mSkewness = 28.42167 chi2(120) = 491.147 Prob>chi2 = 0.0000

Mardia mKurtosis = 119.1498 chi2(1) = 239.486 Prob>chi2 = 0.0000

Henze-Zirkler = 3.806584 chi2(1) = 2707.749 Prob>chi2 = 0.0000

Doornik-Hansen chi2(16) = 75.323 Prob>chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

Test that all means are 0

Hotelling T2 = 305.22

Hotelling F(8,92) = 35.46

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias es, estadísticamente, distinto del vector nulo.

Test that all means are the same

Hotelling T2 = 138.64

Hotelling F(7,93) = 18.60

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 169.53

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 93.63

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que la matriz de datos no es diagonal.

**(b)** *Efectuar similares contrastes, pero considerando cada partición de X por separado.*

Partición 1:

Test for multivariate normality

Mardia mSkewness = 19.60508 chi2(120) = 142.825 Prob>chi2 = 0.0761

Mardia mKurtosis = 78.2123 chi2(1) = 0.200 Prob>chi2 = 0.6549

Henze-Zirkler = .9268645 chi2(1) = 0.015 Prob>chi2 = 0.9039

Doornik-Hansen chi2(16) = 11.505 Prob>chi2 = 0.7773

Por lo tanto, con un nivel de significancia del 10% (excepto con el test de Mardia mSkewness), estos datos no aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

Test that all means are 0

Hotelling T2 = 791.94

Hotelling F(8,32) = 81.22

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias es, estadísticamente, distinto del vector nulo.

Test that all means are the same

Hotelling T2 = 247.35

Hotelling F(7,33) = 29.90

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 156.40

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 30.87

Prob > chi2 = 0.3231

Por lo tanto, con un nivel de significancia del 10%, estos datos no aportan evidencia suficiente para indicar que la matriz de datos no es diagonal.

Partición 2:

Test for multivariate normality

Mardia mSkewness = 13.51716 chi2(120) = 120.983 Prob>chi2 = 0.4577

Mardia mKurtosis = 75.87215 chi2(1) = 1.331 Prob>chi2 = 0.2486

Henze-Zirkler = .9684794 chi2(1) = 1.195 Prob>chi2 = 0.2744

Doornik-Hansen chi2(16) = 22.179 Prob>chi2 = 0.1375

Por lo tanto, con un nivel de significancia del 10%, estos datos no aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

Test that all means are 0

Hotelling T2 = 715.13

Hotelling F(8,42) = 76.62

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias es, estadísticamente, distinto del vector nulo.

Test that all means are the same

Hotelling T2 = 96.28

Hotelling F(7,43) = 12.07

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 68.67

Prob > chi2 = 0.0006

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 34.75

Prob > chi2 = 0.1771

Por lo tanto, con un nivel de significancia del 10%, estos datos no aportan evidencia suficiente para indicar que la matriz de datos no es diagonal.

Partición 3:

Test for multivariate normality

Mardia mSkewness = 51.66604 chi2(120) = 119.165 Prob>chi2 = 0.5044

Mardia mKurtosis = 65.33476 chi2(1) = 3.360 Prob>chi2 = 0.0668

Henze-Zirkler = .9390528 chi2(1) = 2.155 Prob>chi2 = 0.1421

Doornik-Hansen chi2(16) = 27.724 Prob>chi2 = 0.0341

Test that all means are 0

Hotelling T2 = 1555.29

Hotelling F(8,2) = 43.20

Prob > F = 0.0228

Por lo tanto, con un nivel de significancia del 5%, estos datos aportan evidencia suficiente para indicar que el vector de medias es, estadísticamente, distinto del vector nulo.

Test that all means are the same

Hotelling T2 = 44.12

Hotelling F(7,3) = 2.10

Prob > F = 0.2911

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 53.09

Prob > chi2 = 0.0256

Por lo tanto, con un nivel de significancia del 5%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 38.75

Prob > chi2 = 0.0850

Por lo tanto, con un nivel de significancia del 10%, estos datos aportan evidencia suficiente para indicar que la matriz de datos es diagonal.

**(c)** *Proponer soluciones al problema observado, incluso considerando la eventualidad de que existan valores atípicos en la muestra que conlleven al mismo. Aplicar la/s soluciones sugeridas y comentar, brevemente, el resultado encontrado.*

Opción 1:

Test for multivariate normality

Mardia mSkewness = 32.44565 chi2(120) = 101.569 Prob>chi2 = 0.8875

Mardia mKurtosis = 66.93737 chi2(1) = 3.999 Prob>chi2 = 0.0455

Henze-Zirkler = .9306905 chi2(1) = 0.793 Prob>chi2 = 0.3731

Doornik-Hansen chi2(16) = 16.827 Prob>chi2 = 0.3969

Por lo tanto, con un nivel de significancia del 10% (excepto con el test Mardia mKurtosis), estos datos no aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

Test that all means are 0

Hotelling T2 = 235.39

Hotelling F(8,7) = 14.71

Prob > F = 0.0010

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias es, estadísticamente, distinto del vector nulo.

Test that all means are the same

Hotelling T2 = 83.54

Hotelling F(7,8) = 6.82

Prob > F = 0.0074

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 83.43

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 59.67

Prob > chi2 = 0.0004

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que la matriz de datos no es diagonal.

Opción 2:

Test for multivariate normality

Mardia mSkewness = 5.826589 chi2(120) = 100.688 Prob>chi2 = 0.8992

Mardia mKurtosis = 76.76692 chi2(1) = 1.633 Prob>chi2 = 0.2013

Henze-Zirkler = .9373537 chi2(1) = 0.350 Prob>chi2 = 0.5544

Doornik-Hansen chi2(16) = 9.871 Prob>chi2 = 0.8733

Por lo tanto, con un nivel de significancia del 10%, estos datos no aportan evidencia suficiente para indicar que no tienen una distribución normal multivariada.

Test that all means are 0

Hotelling T2 = 311.47

Hotelling F(8,92) = 36.18

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que el vector de medias es, estadísticamente, distinto del vector nulo.

Test that all means are the same

Hotelling T2 = 83.53

Hotelling F(7,93) = 11.21

Prob > F = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las medias de las variables que componen la matriz de datos no son iguales.

Test that covariance matrix is spherical

Adjusted LR chi2(35) = 215.27

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los datos no son esféricos.

Test that covariance matrix is diagonal

Adjusted LR chi2(28) = 150.13

Prob > chi2 = 0.0000

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que la matriz de datos no es diagonal.