



UNIVERSIDAD NACIONAL DE COLOMBIA

ESPECIALIZACIÓN EN CIENCIAS ESTADÍSTICA

DEPARTAMENTO DE ESTADÍSTICA

FACULTAD DE CIENCIAS

— MÉTODOS MULTIVARIADOS APLICADOS —

Teoría Sobre Datos Normales Multivariados
tarea III

Integrantes:

Hugo Hernán Rodríguez Mesa C.C. 1022952094

Medellín, Colombia

Medellin, Marzo 09 de 2024

Contents

1 Ejercicio 01	1
1.1 Resúmenes descriptivos	1
1.2 Gráficos bi-variados	3
1.3 Validación de normalidad univariada y bi-variada	4
2 Ejercicio 02	4
2.1 Resúmenes descriptivos	5
2.2 Gráficos bi-variados	6
2.3 Validación de normalidad univariada y bi-variada	7
2.4 Transformaciones Multivariado	8
Bibliografía	8

1 Ejercicio 01

Generaremos de datos normales multivariados de forma aleatoria con distribución normal multivariada usando la función `mvrnorm` de la librería MASS. ver doc package MASS (2013)

Donde requiere los siguientes argumentos:

- n (el tamaño de la muestra que se quiere generar) generaremos 5000 filas.
- μ (el vector de medias) el cual serca un vector = $[0,0]$.
- Σ la matriz de var-cov $[[1,0.8],[0.8,1]]$.

Usando la funcion `mvnnorm` de la siguiente forma:

```
mvrnorm(n.filas,vector.medias,matriz.covarianza)
```

Obtenemos los siguientes datos normales multivariados, mosramos el head de lso datos.

Table 1: Head Datos aleatorios

X1	X2
-0.0946632	-0.1618326
1.0575654	-0.0650486
2.6863624	2.5877359
2.1464457	1.2873646
-0.4577186	-0.2200488
0.6628962	0.6064511

1.1 Resúmenes descriptivos

1.1.1 Vector de medias

Table 2: Tabla vector de medias

Variables	Mean
X1	-0.0140250130503964
X2	-0.0111704879997269

1.1.2 Matriz de var-cov

Table 3: Tabla var-cov

	X1	X2
X1	1.0105429	0.8026118
X2	0.8026118	1.0032753

1.1.3 Matriz de correlación

Table 4: Tabla correlación

	X1	X2
X1	1.0000000	0.7971098
X2	0.7971098	1.0000000

1.1.4 summary

Table 5: Summary

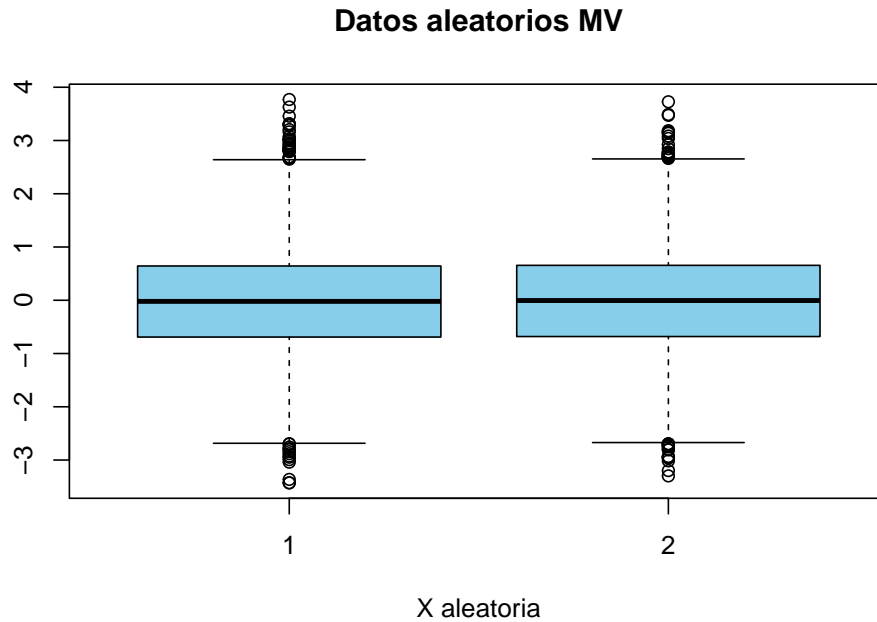
V1	V2
Min. :-3.43040	Min. :-3.298360
1st Qu.: -0.69054	1st Qu.: -0.681497
Median :-0.02103	Median :-0.005251
Mean :-0.01402	Mean :-0.011170
3rd Qu.: 0.64361	3rd Qu.: 0.654482
Max. : 3.76946	Max. : 3.726572

Las variables V1 y V2 presentan distribuciones estadísticas similares. Para V1, la mayoría de los datos se concentran alrededor de la media (-0.02242), con valores que varían desde -3.91040 hasta 3.81811. En cuanto a V2, los datos tienen una dispersión similar en torno a la media (-0.004921), con valores que oscilan entre -3.453670 y 3.731694.

En cuanto a la relación entre las variables, la matriz de correlación muestra una fuerte correlación positiva de 0.8029123 entre V1 y V2. La matriz de covarianza refleja una covarianza positiva de 0.7999135 entre ambas variables, indicando una tendencia similar en sus variaciones.

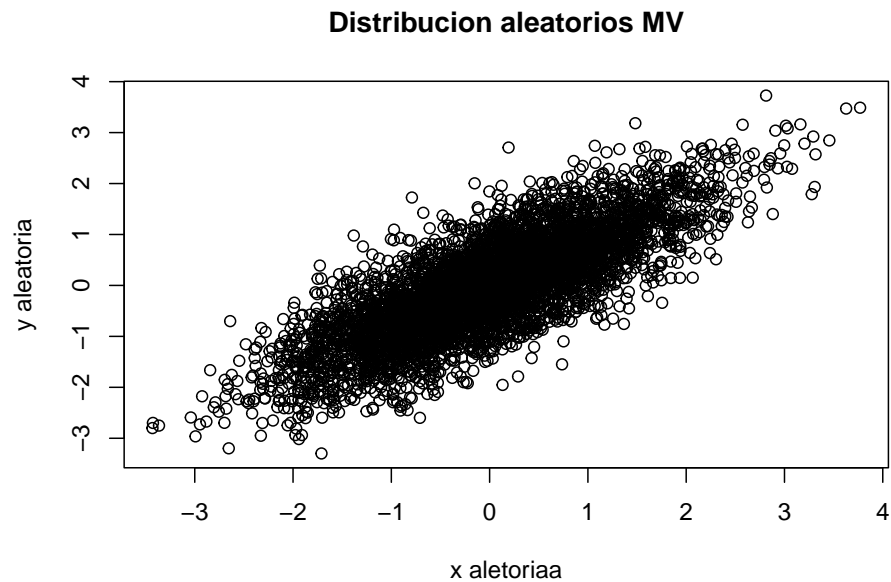
1.2 Gráficos bi-variados

1.2.1 boxplot

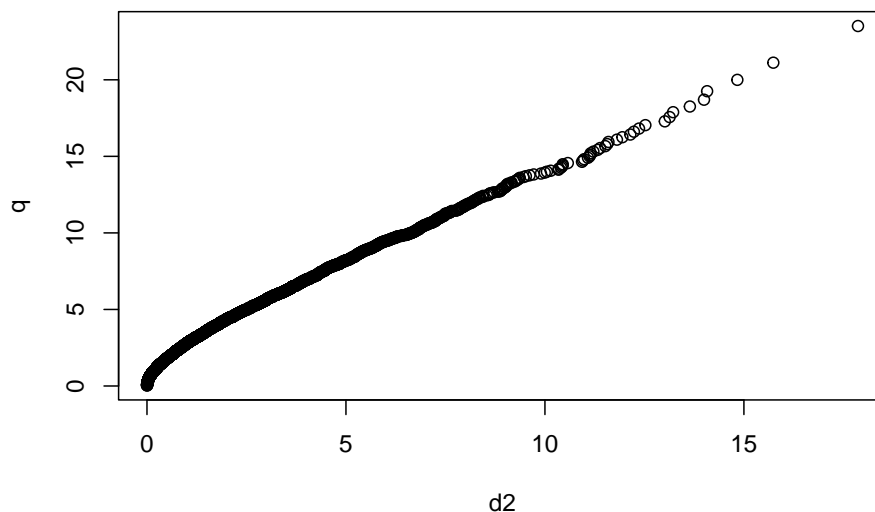


1.2.2 Plot

De acuerdo con el grafico de dispersion se observa una relacion lineal positiva entre las variables aleatorias x_1 , x_2 , de hecho en la tabla de correlacion es cerca al 80% aproximadamente. De manera que hay un alto nivel de dependencia entre las variables.



1.3 Validación de normalidad univariada y bi-variada



Salidas básicas prueba normalidad con la función `mvn` de tipo mardia y univariada Shapiro y Wilk.

Table 6: Multivariate Normality

multivariateNormality.Test	multivariateNormality.Statistic	multivariateNormality.p.value	multivariateNormality.Result
Mardia Skewness	7.7453044905211	0.101366375282353	YES
Mardia Kurtosis	1.05965480762196	0.289301669625988	YES
MVN	NA	NA	YES

Table 7: Univariate Normality

univariateNormality.Test	univariateNormality.Variable	univariateNormality.Statistic	univariateNormality.p.value
Shapiro-Wilk	Column1	0.9993	0.0387
Shapiro-Wilk	Column2	0.9996	0.4541

Parece que gran cantidad de los puntos cae sobre una recta de pendiente 1, poodria aprobarse la prueba de normalidad.

2 Ejercicio 02

Los datos utilizados en este análisis descriptivo fueron extraídos de la página oficial del examen Saber 11, correspondientes al periodo 2015-2, los cuales se pueden descargar desde el sitio web: Datos-Icfes. Esta tarea Se limita a trabajar únicamente al departamento de Cesar, Colombia y los promedios de sociales y matematicas.

Table 8: Head Datos pruebas Cesar

PROMMATEMATICA	PROMSOCIALESYCIUDADANAS
55.36	53.62
49.85	49.11
48.49	45.49
53.17	51.86
49.68	49.63
59.88	59.18

2.1 Resúmenes descriptivos

2.1.1 Vector de medias

Table 9: Vector de medias

	Mean
PROMMATEMATICA	46.85413
PROMSOCIALESYCIUDADANAS	46.53811

2.1.2 Matriz de var-cov

Table 10: Tabla de var-cov

	PROMMATEMATICA	PROMSOCIALESYCIUDADANAS
PROMMATEMATICA	32.24538	26.84052
PROMSOCIALESYCIUDADANAS	26.84052	28.69088

2.1.3 Matriz de correlación

Table 11: Tabla correlación

	PROMMATEMATICA	PROMSOCIALESYCIUDADANAS
PROMMATEMATICA	1.0000000	0.8824403
PROMSOCIALESYCIUDADANAS	0.8824403	1.0000000

2.1.4 summary

Table 12: Summary

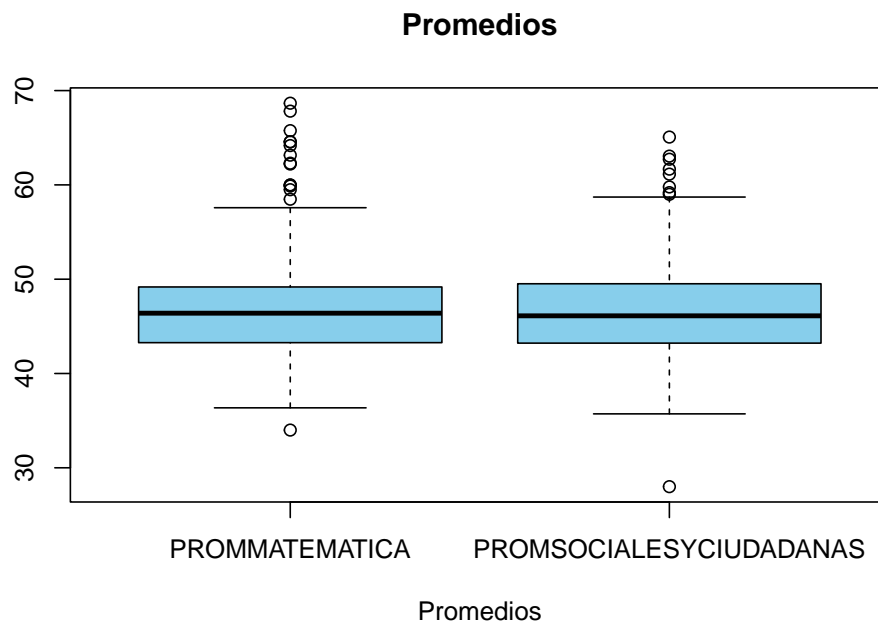
--

PROMMATEMATICA	PROMSOCIALESYCIUDADANAS
Min. :34.00	Min. :28.00
1st Qu.:43.27	1st Qu.:43.23
Median :46.40	Median :46.12
Mean :46.85	Mean :46.54
3rd Qu.:49.17	3rd Qu.:49.48
Max. :68.66	Max. :65.07

Las calificaciones promedio en PROMMATEMATICA oscilan entre 34.00 y 68.66 (media: 46.85), y en PROMSOCIALESYCIUDADANAS entre 28.00 y 65.07 (media: 46.54). Existe una fuerte correlación positiva (0.88) entre ambas materias, indicando que los estudiantes tienden a tener rendimientos similares.

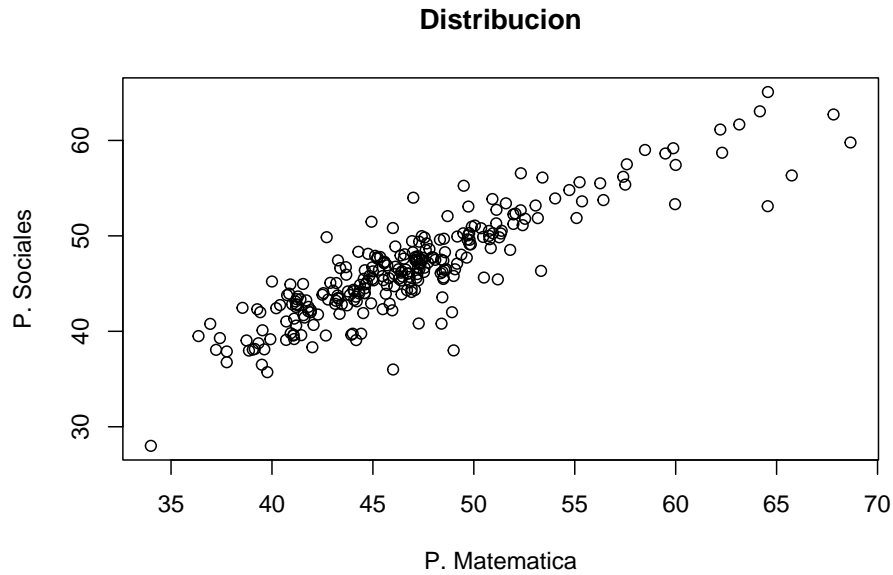
2.2 Gráficos bi-variados

2.2.1 boxplot

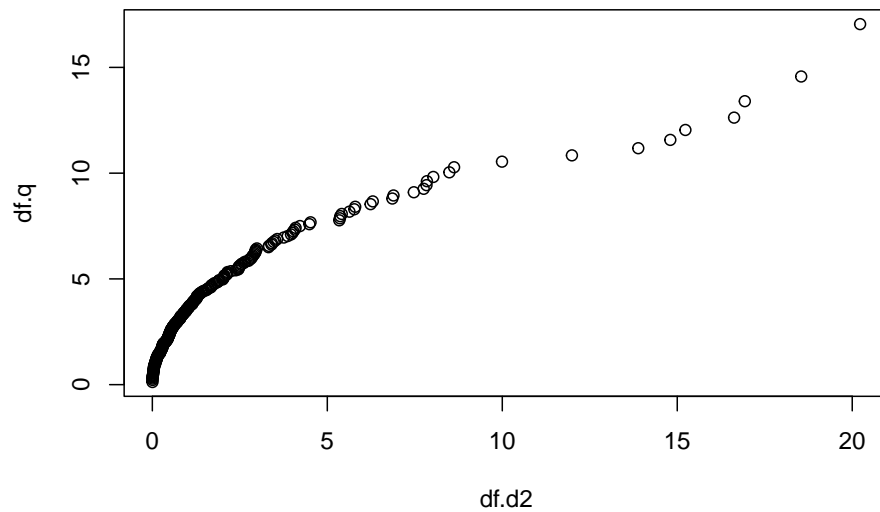


2.2.2 Plot

De acuerdo con el grafico de dispersion se observa una relacion lineal positiva entre las variables aleatorias x_1 , x_2 , de hecho en la tabla de correlacion es cerca al 80% aproximadamente. De manera que hay un alto nivel de dependencia entre las variables.



2.3 Validación de normalidad univariada y bi-variada



Salidas básicas prueba normalidad con la función `mvn` de tipo royston y univariada Shapiro y Wilk.

Table 13: Multivariate Normality

<code>multivariateNormality.Test</code>	<code>multivariateNormality.H</code>	<code>multivariateNormality.p.value</code>	<code>multivariateNormality.Result</code>
Royston	43.40335	0	NO

Table 14: Univariate Normality

<code>univariateNormality.Test</code>	<code>univariateNormality.Variable</code>	<code>univariateNormality.Statistic</code>	<code>univariateNormality.p.value</code>
Shapiro-Wilk	PROMMATEMATICA	0.9271	<0.001

Parece quede los puntos NO caen sobre una recta de pendiente 1, entonces no poodria aprobarse la prueba de normalidad.

2.4 Transformaciones Multivariado

Para cada una de las variables. Sea λ_1, λ_2 las transformaciones de potencia para las p variables Promedio matematicas y Promedio sociales, y transformaremos los datos con `bcnPower`.

Table 15: Tabla nuevos datos normalizados

PROMMATEMATICA ^{-1.71}	PROMSOCIALESYCIUDADANAS ^{-0.68}
0.0010405	0.0669714
0.0012450	0.0710879
0.0013053	0.0748813
0.0011150	0.0685063
0.0012523	0.0705814
0.0009098	0.0626323

2.4.1 Nueva prueba de normalidad

Salidas básicas prueba normalidad con la funcion `mvn` de tipo royston y univariada Shapiro y Wilk.

Table 16: Multivariate Normality

multivariateNormality.Test	multivariateNormality.H	multivariateNormality.p.value	multivariateNormality.Result
Royston	16.43209	0.0001443	NO

Table 17: Univariate Normality

univariateNormality.Test	univariateNormality.Variable	univariateNormality.Statistic	univariateNormality.p.value	univariateNormality.Normality
Shapiro-Wilk	PROMMATEMATICA ^{-1.71}	0.9894	0.0511	YES
Shapiro-Wilk	PROMSOCIALESYCIUDADANAS ^{-0.68}	0.9755	0.0002	NO

Bibliografía

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D. & Ripley, M.B. (2013). Package ‘mass’. *Cran r*, **538**, 113–120.