[Capítulo IV 275](#_Toc514131370)

[4.- Análisis Multivariado 275](#_Toc514131371)

[4.1 Introducción 275](#_Toc514131372)

[4.2 Definiciones 276](#_Toc514131373)

[4.3 Análisis de la matriz de correlación 280](#_Toc514131374)

[4.3.1 Análisis de los mayores coeficientes de correlación 280](#_Toc514131375)

[4.3.2 Análisis de los menores coeficientes de correlación 285](#_Toc514131376)

# Capítulo IV

# 4.- Análisis Multivariado

## 4.1 Introducción

El análisis multivariado está constituido por un conjunto de técnicas estadísticas diseñadas para extraer simultáneamente información de un grupo de variables aleatorias. En particular las técnicas multivariadas que se aplicarán en este capítulo son cinco: correlación lineal, análisis de varianza, análisis de componentes principales, correlación canónica y análisis de contingencia.

## 4.2 Definiciones

Para definir matemáticamente los métodos multivariados que se utilizarán en este capítulo, es necesario partir de definición de la matriz de datos. La información de las variables debe ser representada en forma de un arreglo rectangular de p columnas por n filas denotado por **X**, donde cada columna es un vector en Rn y cada fila es un vector en Rp; el valor p corresponde al número de variables aleatorias (X1, X2, ... ,Xp) y el valor n al número de unidades de investigación, entonces a la matriz **X** se la denomina matriz de datos.



La matriz **X** contiene todas las observaciones de todas las variables, el valor de Xjk representa la j-ésima observación de la k-ésima variable aleatoria, para j=1,2,...,n y k=1,2,...,p.

Sea **X**T = (X1, X2, ... ,Xp), un vector aleatorio, se define el vector de medias **μ**, correspondientes al vector **X** como sigue,

Donde el valor μk representa el valor esperado de la variable aleatoria Xk, es decir μk = E [Xk], para k =1, 2, ... , p.

Luego de definir el vector de medias **μ**, se define la matriz de varianzas y covarianzas **Σ**, como se muestra a continuación,

Sea σij = E ( Xi - μi ) ( Xj - μj ), para i, j = 1,2,...,p, σij representa la covarianza entre Xi y Xj, entonces **Σ** = E (**X - μ**) (**X - μ**)T, lo cual se puede expresar como







Cuando i=j se tiene que σij =σi2, que es la varianza de la variable aleatoria Xi. La matriz **ρ** muestra los valores de ρij que representan los coeficientes de correlación entre las variables Xi y Xj,



El coeficiente indica si existe o no, relación lineal entre estas dos variables, y se obtiene con



donde σi y σj son las desviaciones estándar de Xi y Xj respectivamente, cuando i=j el coeficiente de correlación es igual a 1, además este coeficiente se encuentra en el intervalo (–1,1).

**Prueba de que -1 ≤ ρij**

Sabemos que la varianza de cualquier variable aleatoria es mayor a cero, Var ( X) ≥ 0, entonces

**Prueba de que ρij ≤ 1**

Sabemos que la varianza de cualquier variable aleatoria es mayor a cero, Var ( X) ≥ 0, entonces

Por lo tanto se cumple que -1 ≤ ρij ≤ 1.

## 4.3 Análisis de la matriz de correlación

De los resultados que se obtuvieron de la matriz de correlación, de las sesenta variables de estudio utilizadas para medir el conocimiento en matemáticas y lenguaje, de los estudiantes de décimo año de educación básica, de los colegios fiscales rurales del cantón Guayaquil, en esta sección se analizarán los mayores y menores coeficientes de correlación entre las variables aleatorias estudiadas.

### 4.3.1 Análisis de los mayores coeficientes de correlación

Para realizar este análisis se consideró como significativos los coeficientes de correlación que en valor absoluto fuesen mayores a 0.6. La matriz de correlación obtenida, de todas las variables de estudio se muestra en el anexo número dos. En éste análisis se determinó que existen tres pares de variables aleatorias correlacionadas entres sí.

* **Variable identificación de diptongos con la variable identificación de triptongos**

Se determinó que las variables observables X34 identificación de diptongos y X35 identificación de triptongos, son linealmente dependientes, pues el coeficiente de correlación lineal entre estas variables es 0.831.

 La matriz de correlación de las variables X34 y X35 es:

|  |  |  |
| --- | --- | --- |
|  | X34 | X35 |
| X34 | 1 | 0.833 |
| X35 | 0.833 | 1 |

**Gráfico 4.1**

**Diagrama de dispersión de las variables identificación de diptongos e identificación de triptongos**



La dispersión entre las variables X34 y X35 se puede observar en el gráfico 4.1, en el cual se ilustra el número de observaciones en cada uno de los puntos graficados, la relación lineal entre estas dos variables es directa, es decir que a medida que los valores de la variable X34 aumentan, los valores de X35 también aumentan. La variable aleatoria identificación de diptongos puede tomar valores de 0 a 10 los cuales representan el número de diptongos que el estudiante identificó correctamente, mientras que la variable aleatoria identificación de triptongos puede tomar valores de 0 a 5. Los puntos graficados indican que existen estudiantes que identificaron correctamente el número de diptongos representados por el valor de la accisa y el número de triptongos representados por el valor de la ordenada. En el caso contrario, por ejemplo la coordenada (4,0) que no está graficada indica que ningún estudiante identificó correctamente 4 diptongos y ningún triptongo.

* **Variable identificación del sujeto de la oración con la variable identificación del predicado de la oración**

Las variables X50 identificación del sujeto de la oración y X51 identificación del predicado de la oración, son linealmente dependientes, el coeficiente de correlación lineal entre estas variables es de 0.665. A pesar que éste valor es mayor a 0.6, la dependencia lineal entre las dos variables, no es fuerte.

La matriz de correlación de las variables X50 y X51 es:

|  |  |  |
| --- | --- | --- |
|  | X50 | X51 |
| X50 | 1 | 0.665 |
| X51 | 0.665 | 1 |

La dependencia lineal existente entre las variables X50 y X51 determinadas por el coeficiente de correlación, el cual es mayor a 0.6, no se puede apreciar claramente en el gráfico 4.2, donde se muestra la dispersión entre estas dos variables, debido a que el valor de este coeficiente no es muy cercano a uno.

**Gráfico 4.2**

**Diagrama de dispersión de las variables identificación del sujeto de la oración e identificación del predicado de la oración**



Tanto la variable aleatoria identificación del sujeto como la variable aleatoria identificación del predicado pueden tomar valores entre 0 y 4, los cuales representan el número de sujetos o predicados identificados correctamente. Como se puede observar en el gráfico 4.2 a excepción de los puntos con coordenadas (0,3) y (1,0), todos los demás resultados posibles de respuestas se presentan entre esta dos variables aleatorias, con el respectivo número de observaciones.

* **Variable multiplicación de números racionales con la variable división de números racionales**

Existe dependencia lineal entre las variables X8 multiplicación de fracciones y X10 división de números racionales, el coeficiente de correlación entre ambas variables es 0.627.

La matriz de correlación de las variables X8 y X10 es:

|  |  |  |
| --- | --- | --- |
|  | X8 | X10 |
| X8 | 1 | 0.627 |
| X10 | 0.627 | 1 |

**Gráfico 4.3**

**Diagrama de dispersión entre las variables multiplicación de números racionales y división de números racionales**



Los valores que pueden tomar las observaciones son 0 (respuesta incorrecta) o 1 (respuesta correcta) para ambas variables aleatorias, en el gráfico 4.3 se muestra el diagrama de dispersión entre las variables X8 y X10. En este gráfico se puede observar que no existe el punto con coordenadas (1,0), lo que significa que ningún estudiante entrevistado realizó correctamente la multiplicación de números racionales e incorrectamente la división de números racionales.

### 4.3.2 Análisis de los menores coeficientes de correlación

Cuando dos variables aleatorias X y Y son independientes, entonces la covarianza y el coeficiente de correlación entre estas variables, es igual a cero; sin embargo, si el valor de la covarianza y el coeficiente de correlación entre estas dos variables es igual a cero, no implica, que las variables aleatorias son independientes, sino que, no existe dependencia lineal entre ambas variables.

Otros resultados obtenidos a partir del análisis de la matriz de correlación, fueron los de los pares de variables aleatorias menos correlacionadas, para lo cual se consideraron los coeficientes de correlación lineal menores a 0.000, a continuación se detallan los tres pares de variables aleatorias que cumplen con esta condición.

* **Variable identificación de palabras agudas con la variable identificación de palabras diminutivas**

El coeficiente de correlación entre las variables observables X37 identificación de palabras agudas y X45 identificación de palabras diminutivas es 0.000, el cual indica que esta variables aleatorias no son linealmente dependientes. La matriz de correlación de las variables X37 y X45 es:

La matriz de correlación de las variables X37 y X45 es:

|  |  |  |
| --- | --- | --- |
|   | X37 | X45 |
| X37 | 1 | 0 |
| X45 | 0 | 1 |

La dispersión entre las variables X37 y X45 se puede observar en el gráfico 4.4. La variable aleatoria identificación de palabras agudas puede tomar valores de 0 a 5 los cuales representan el número de palabras agudas que el estudiante identificó correctamente, mientras que la variable aleatoria identificación de palabras diminutivas puede tomar valores de 0 a 2. Los puntos graficados indican que existen estudiantes que identificaron correctamente el número de palabras agudas representados por el valor de la accisa y el número de palabras diminutivas representados por el valor de la ordenada. Se puede observar en el gráfico que el mayor número de observaciones corresponden a las coordenadas del punto (0,0), que corresponde a los estudiantes que no identificaron correctamente las palabras agudas y las palabras diminutivas.

**Gráfico 4.4**

**Diagrama de dispersión de las variables identificación de palabras agudas e identificación de palabras diminutivas**



* **Variable número de relaciones de orden correctas con la variable número de palabras definidas correctamente.**

Entre las variables X13 número de relaciones de orden correctas y X48 número de palabras definidas correctamente, el coeficiente de correlación lineal es 0, éste valor indica que las variables aleatorias X13 y X48 no son linealmente dependientes.

La matriz de correlación de las variables X13 y X48 es:

|  |  |  |
| --- | --- | --- |
|  | X13 | X48 |
| X13 | 1 | 0 |
| X48 | 0 | 1 |

La dispersión existente entre las variables X13 y X48, se ilustra claramente en el gráfico 4.5.

**Gráfico 4.5**

**Diagrama de dispersión de las variables número de relaciones de orden correctas y número de palabras definidas correctamente**



Las variables aleatorias número de relaciones de orden correctas y número de palabras definidas correctamente pueden tomar valores entre 0 y 5, y entre 0 y 4, respectivamente. Como se puede observar en el gráfico 4.5 entre los puntos que no están graficados, se encuentra (4,5) que corresponden al máximo número de definiciones de palabras y relaciones de orden correctas, es decir que ningún estudiante entrevistado pudo contestar correctamente las preguntas correspondientes a las variables X13 y X48.

* **Variable definiciones de conjuntos de números con la variable ecuación lineal con una incógnita**

De acuerdo al valor del coeficiente de correlación obtenido entre las variables X11 definiciones de conjuntos de números y X29 ecuación lineal con una incógnita, se determinó que no existe dependencia lineal entre ambas variables aleatorias.

La matriz de correlación de las variables X11 y X29 es:

|  |  |  |
| --- | --- | --- |
|  | X11 | X29 |
| X11 | 1 | 0 |
| X29 | 0 | 1 |

Las variables aleatorias definiciones de conjuntos de números y ecuación lineal con una incógnita, pueden tomar los siguientes valores, para la primera variable de 0 a 4, dependiendo del número de respuestas correctas que obtuvo el estudiante y para la segunda variable 0 si la respuesta es incorrecta o 1 si es correcta, en el gráfico 4.6 se muestra el diagrama de dispersión entre las variables X11 y X29. En este gráfico se puede observar claramente que ningún estudiante entrevistado respondió correctamente la pregunta correspondiente a la variable aleatoria X11 o a la pregunta correspondiente a la variable X29.

**Gráfico 4.6**

**Diagrama de dispersión entre las variables definiciones de conjuntos de números y ecuación lineal con una incógnita**

