[4.4 Análisis de contingencia 291](#_Toc514133129)

[4.4.1 Análisis de contingencia entre las variables edad y calificación de lenguaje 295](#_Toc514133130)

[4.4.2 Pruebas entre las variables sexo y calificación de lenguaje 297](#_Toc514133131)

[4.4.3 Pruebas entre las variables ubicación del establecimiento educativo y calificación de lenguaje 298](#_Toc514133133)

[4.4.4 Pruebas entre las variables jornada de trabajo del establecimiento educativo y calificación de lenguaje 300](#_Toc514133135)

[4.4.5 Pruebas entre las variables edad y calificación de matemáticas 302](#_Toc514133137)

[4.4.6 Pruebas entre las variables sexo y calificación de matemáticas 303](#_Toc514133138)

[4.4.7 Pruebas entre las variables ubicación del establecimiento educativo y calificación de matemáticas 305](#_Toc514133139)

[4.4.8 Pruebas entre las variables jornadas del establecimiento educativo y calificación de matemáticas 307](#_Toc514133140)

[4.4.10 Resultados del análisis de contingencia entre otras variables aleatorias 310](#_Toc514133141)

## 4.4 Análisis de contingencia

En el análisis de la matriz de correlación de las variables aleatorias que se utilizaron 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, se detectaron en base a los coeficientes de correlación lineal, aquellos pares de variables aleatorias que son independientes ó linealmente dependientes; mientras que el análisis de contingencia se realiza con el objetivo de determinar la dependencia ó independencia lineal o no, de dos métodos o criterios de clasificación de las variables aleatorias observadas.

El análisis de contingencia se lo define de la siguiente forma, sean X y Y dos variables aleatorias observables y sea n el número de observaciones de cada variable, para cada una de las variables se define una clasificación con las cuales se construye un arreglo matricial de r filas y c columnas, donde c es el número de criterios de clasificación de la variable X y r el número de criterios de clasificación de la variable Y, este arreglo matricial se denomina tabla de contingencia.

En las tablas de contingencia se contabilizan las observaciones que cumplen simultáneamente con cada criterio de clasificación de cada variable. Deben de haber al menos dos criterios de clasificación por variable los cuales deben ser exhaustivos y mutuamente excluyentes.

Para realizar el análisis de contingencia se construye un contraste de hipótesis para probar la hipótesis nula de que los criterios de clasificación de la variable aleatoria X son independientes de los criterios de clasificación de la variable aleatoria Y.

A continuación se muestra un modelo de tabla de contingencia en el que se puede observar como se contabilizan las observaciones de acuerdo a los criterios de clasificación de cada una de las variables aleatorias.

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación de la variable aleatoria Y** |  |
| Criterios de clasificación de la variable aleatoria X |  1 | **2** | ... | **c** | **Total** |
| **1** | n11E [ n11] | n12E [ n12] | ... | n1cE [ n1c] | **r1** |
| **2** | n21E [ n21] | n22E [ n22] | ... | n2cE [ n2c] | **r2** |
| **…** | **…** | **…** |  | **…** | **…** |
| **r** | nr1E [ nr1] | nr2E [ nr2] | ... | nrc E [ nrc] | **rr** |
| **Total** | **c1** | **c2** | **...** | **cc** | **n** |

Los valores de nij representan el número de observaciones que cumplen con el criterio de clasificación i, para i=1,2,...,r de la variable aleatoria X y con el criterio de clasificación j, para j=1,2,...,c de la variable aleatoria Y.

Los valores de ri representan el total de observaciones que están sujetas al criterio de observación ide la variable aleatoria X, análogamente los valores de cj representan el total de observaciones que caen en el criterio de observación j de la variable aleatoria Y. El valor de n representa el total de observaciones de la muestra o de la población.

El valor esperado de nij denotado por E[ nij ] es igual al producto del total de observaciones del renglón i (ri) por el total de observaciones de la columna j (cj)dividido entre el total de observaciones (n).

E[ nij ] = ( ri \* cj ) / n.

El contraste de hipótesis que se construye para probar la independencia de los criterios de clasificación de las variables aleatorias X y Y es el siguiente:

H0 : Los criterios de clasificación son independientes

vs.

Ha : Los criterios de clasificación son dependientes

El estadístico de prueba que utiliza la información contenida en la tabla de contingencias, es una variable aleatoria ji-cuadrado denotada por χ2 con (r-1)(c-1) grados de libertad y se lo calcula como se muestra a continuación:

Estadístico de prueba: 

donde, 

Entonces se rechaza H0 en favor de Ha, con (1-α)100%, si:



Se determina valor p de la prueba o nivel de significancia alcanzado, que es el mínimo valor de α para el cual los datos observados indican que se tendría que rechazar la hipótesis nula.

A continuación se realizará análisis de contingencia entre las variables generales (edad, sexo, ubicación y sección) y las variables de calificación de lenguaje y calificación de matemáticas, utilizando el procedimiento descrito en esta sección.

### 4.4.1 Análisis de contingencia entre las variables edad y calificación de lenguaje

El número de criterios de clasificación que se utilizará para cada variable es dos, debido a que al realizar la prueba con más criterios existían celdas de la tabla de contingencia que contienen menos de cinco observaciones lo cual no es recomendable.

Definimos primeramente los criterios de clasificación de la variable aleatoria edad, el criterio **A** corresponde a las edades entre 10 y 15 años y el criterio **B** contiene a las edades de 16 a 22 años.

#### **Tabla CXXXII**

#### **Criterios de clasificación para las variables edad y calificación de lenguaje**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la calificación de lenguaje** |  |
| **Criterios de clasificación para la edad** | C | **D** | **Total** |
| **A** | 6065.209 | 6155.79 | **121** |
| **B** | 3024.79 | 1621.209 | **46** |
| **Total** | **90** | **77** | **167** |

Para la variable aleatoria calificación de lenguaje el criterio de **C** es el de las calificaciones desde 0 a 44 puntos sobre cien, y el criterio **D** corresponde a las calificaciones entre 45 y 83 puntos. En la tabla CXXXII se muestran las observaciones agrupadas de acuerdo a los criterios de clasificación antes mencionados.

Los valores esperados nij obtenidos son mostrados en la tabla CXXXII, estos valores son utilizados para calcular el estadístico de prueba. A continuación se construye el contraste de hipótesis para determinar la independencia de las variables edad y calificación de lenguaje.

**H0:** Los criterios de clasificación de las variables aleatorias

 edad y calificación de lenguaje son independientes

vs.

**Ha::** Los criterios de clasificación de las variables aleatorias

 edad y calificación de lenguaje son dependientes

Valor del estadístico de prueba χ2= 3.277

Grados de libertad 1

Valor p = 0.070256

En vista de que el valor p, es 0.10 ≤ p ≤ 0.05, simplemente mostramos el mismo.

### 4.4.2 Pruebas entre las variables sexo y calificación de lenguaje

Para realizar el análisis de contingencia entre estas dos variables se definieron los siguientes criterios de clasificación. Para la variable sexo el criterio **A** corresponde al sexo femenino y el criterio **B** al sexo masculino; y los criterio para la variable aleatoria calificación de lenguaje son **C** que corresponde a las calificaciones entre 0 y 44 puntos, y el criterio **D** que contiene las calificaciones de 45 a 83 puntos. Las observaciones que cumplen con estos criterios de clasificación son mostradas en la tabla CXXXIII.

#### **Tabla CXXXIII**

#### **Criterios de clasificación para las variables sexo y calificación de lenguaje**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de lenguaje** |  |
| **Criterios de clasificación para la variable sexo** | C | **D** | **Total** |
| **A** | 4546.317 | 4038.682 | **85** |
| **B** | 4644.682 | 3637.317 | **82** |
| **Total** | **91** | **76** | **167** |

Para determinar la independencia de los criterios de clasificación de las variables sexo y calificación de lenguaje se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables sexo y

 calificación de lenguaje son independientes

vs.

**Ha**: Los criterios de clasificación de las variables sexo y

 calificación de lenguaje son dependientes

Valor del estadístico de prueba χ2= 3.84146

Grados de libertad 1

Valor p = 0.16767

# Existe evidencia estadística para no rechazar H0, por lo tanto se concluye que los criterios de clasificación son independientes.

### 4.4.3 Pruebas entre las variables ubicación del establecimiento educativo y calificación de lenguaje

Los criterios de clasificación utilizados para el análisis de contingencia de estas dos variables son para la variable ubicación del establecimiento educativo el criterio **A** que corresponde a las zonas de Progreso, Sabana grande y Puná , y el criterio **B** contiene a las zonas de Posorja y Tenguel.

En lo que respecta a la variable calificación de lenguaje los criterios de clasificación son: **C** que contiene a las calificaciones de 0 a 44 puntos, y el criterio **D** que corresponde a las calificaciones entre 45 y 83 puntos. Los totales de las observaciones que cumplen con cada uno de estos criterios son mostrados en la tabla CXXXIV.

#### **Tabla CXXXIV**

####  **Criterios de clasificación para las variables ubicación y calificación de lenguaje**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la calificación de lenguaje** |  |
| **Criterios de clasificación para la ubicación** | C | **D** | **Total** |
| **A** | 1717.784 | 3736.215 | **54** |
| **B** | 3837.215 | 7575.784 | **113** |
| **Total** | **55** | **112** | **167** |

Para determinar la independencia de las variables ubicación del establecimiento educativo y calificación de lenguaje se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables ubicación

y calificación de lenguaje son independientes

vs.

**Ha**: Los criterios de clasificación de las variables ubicación

y calificación de lenguaje son dependientes

Valor del estadístico de prueba χ2= 3.84146

Grados de libertad 1

Valor p = 0.78245

# Existe evidencia estadística para no rechazar H0, por lo tanto se concluye que los criterios de clasificación son independientes.

### 4.4.4 Pruebas entre las variables jornada de trabajo del establecimiento educativo y calificación de lenguaje

Los criterios de clasificación definidos para las variable analizadas en esta sección son para la variable jornada el criterio **A** corresponde a la jornada matutina y el criterio **B** contiene a las jornadas vespertina y nocturna.

Para la variable calificación de lenguaje los criterios son **C** que corresponde a las calificaciones entre 0 y 44 puntos; y el criterio **D** que contiene a las calificaciones desde 45 a 83. En la tabla CXXXV se pueden observar los totales de las observaciones que cumplen con cada criterio de clasificación.

#### **Tabla CXXXV**

####  **Criterios de clasificación para las variables jornada y calificación de lenguaje**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de lenguaje** |  |
| **Criterios de clasificación para la variable jornada** | C | **D** | **Total** |
| **A** | 7978.467 | 1212.539 | **101** |
| **B** | 6565.532 | 1110.467 | **66** |
| **Total** | **144** | **23** | **167** |

Para determinar la independencia de los criterios de clasificación definidos, de las variables jornada de trabajo del establecimiento educativo y calificación de lenguaje se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables jornada

y calificación de lenguaje son independientes

vs.

**Ha**: Los criterios de clasificación de las variables jornada

y calificación de lenguaje son dependientes

Valor del estadístico de prueba χ2= 0.057749

Grados de libertad 1

Valor p = 0.81008

# Existe evidencia estadística para no rechazar H0, por lo tanto se concluye que los criterios de clasificación son independientes.

### 4.4.5 Pruebas entre las variables edad y calificación de matemáticas

Para realizar el análisis de contingencia con estas dos variables se definieron los criterios de clasificación de la siguiente manera, para la variable edad se utilizó el mismo criterio del análisis de la sección 4.5.1, es decir el criterio **A** corresponde a las edades de 10 a 15 años y el criterio **B** contiene a las edades de 16 a 22 años.

#### **Tabla CXXXVI**

#### **Criterios de clasificación para las variables edad y calificación de matemáticas**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de matemáticas** |  |
| **Criterios de clasificación para la variable edad** | C | **D** | **Total** |
| **A** | 7673.175 | 4547.82 | **121** |
| **B** | 2527.82 | 2118.178 | **46** |
| **Total** | **101** | **66** | **167** |

En lo que respecta a la variable calificación de matemáticas el criterio de **C** contiene a las calificaciones de 0 a 14 puntos, y el criterio **D** corresponde a las calificaciones entre 15 y 32 puntos.

Para determinar la independencia de los criterios de clasificación entre las variables aleatorias edad y calificación de matemáticas se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables edad

y calificación de matemáticas son independientes

vs.

**Ha**: Los criterios de clasificación de las variables edad

y calificación de matemáticas son dependientes

Valor del estadístico de prueba χ2= 0.9985

Grados de libertad 1

Valor p = 0.31767

Existe evidencia estadística para no rechazar H0, por lo tanto se concluye que los criterios de clasificación son independientes.

### 4.4.6 Pruebas entre las variables sexo y calificación de matemáticas

La definición de los criterios de clasificación para las variables analizas en esta sección son, para la variable sexo el criterio **A** corresponde al sexo femenino y el criterio **B** al sexo masculino, y para la variable calificación de matemáticas el criterio de **C** es el de las calificaciones entre 0 y 14 puntos y el criterio **D** corresponde a la calificaciones entre 15 y 32 puntos

#### **Tabla CXXXVII**

####  **Criterios de clasificación para las variables sexo y calificación de matemáticas**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de matemáticas** |  |
| **Criterios de clasificación para la variable sexo** | C | **D** | **Total** |
| **A** | 5152.425 | 3432574 | **85** |
| **B** | 5250.574 | 3031.425 | **82** |
| **Total** | **103** | **64** | **167** |

Para determinar la independencia de los criterios de clasificación de las variables sexo y calificación de matemáticas se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables sexo

y calificación de matemáticas son independientes

vs.

**Ha**:: Los criterios de clasificación de las variables sexo

y calificación de matemáticas son dependientes

Valor del estadístico de prueba χ2= 0.20588

Grados de libertad 1

Valor p = 0.65001

Existe evidencia estadística para no rechazar H0, por lo tanto se concluye que los criterios de clasificación son independientes.

### 4.4.7 Pruebas entre las variables ubicación del establecimiento educativo y calificación de matemáticas

Los criterios de clasificación de las variables que se analizarán en esta sección son para la variable ubicación el criterio **A** corresponde a las zonas de Progreso, Sabana grande y Puná , el criterio **B** contiene a las zonas de Posorja y Tenguel. Para la variable calificación de matemáticas se definieron los criterios **C** que corresponde a las calificaciones entre 0 y 14 puntos y **D** que corresponde a las calificaciones de 15 a 32 puntos.

Una vez definidos los criterios de clasificación de las variables para el análisis de contingencia se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables ubicación

y calificación de matemáticas son independientes

vs.

**Ha**: Los criterios de clasificación de las variables ubicación

y calificación de matemáticas son dependientes

#### **Tabla CXXXVIII**

#### **Criterios de clasificación para las variables ubicación y calificación de matemáticas**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de matemáticas** |  |
| **Criterios de clasificación para la variable ubicación** | C | **D** | **Total** |
| **A** | 2433.592 | 3121.407 | **55** |
| **B** | 7868.407 | 3443.592 | **112** |
| **Total** | **102** | **65** | **167** |

Valor del estadístico de prueba χ2= 0.10494

Grados de libertad 1

Valor p = 0.00119775

Existe evidencia estadística para rechazar H0, en favor de Ha por lo tanto se concluye que los criterios de clasificación son dependientes.

### 4.4.8 Pruebas entre las variables jornadas del establecimiento educativo y calificación de matemáticas

Para realizar el análisis de contingencia entre las variables jornada y calificación de matemáticas se definieron los siguientes criterios de clasificación, para la variable jornada el criterio **A** corresponde a la jornada matutina y el criterio **B** contiene a la jornada vespertina y nocturna; en lo que respecta a la variable calificación de matemáticas el criterio **C** es el de las calificaciones entre 0 y 14 puntos y el criterio **D** corresponde a las calificaciones desde 15 a 32 puntos. En la tabla CXXXIX se muestran los las observaciones que cumplen estos criterios.

#### **Tabla CXXXIX**

#### **Criterios de clasificación para la variable jornada y calificación de matemáticas**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de matemáticas** |  |
| **Criterios de clasificación para la variable jornada** | C | **D** | **Total** |
| **A** | 8387.95 | 6156.047 | **144** |
| **B** | 1914.047 | 48.952 | **23** |
| **Total** | **102** | **65** | **167** |

Para determinar la independencia de los criterios de clasificación de las variables jornada y calificación de matemáticas se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables jornada

y calificación de matemáticas son independientes

vs.

**Ha**: Los criterios de clasificación de las variables jornada

y calificación de matemáticas son dependientes

Valor del estadístico de prueba χ2= 5.2014

Grados de libertad 1

Valor p = 0.022568

Existe evidencia estadística para rechazar H0, en favor de Ha por lo tanto se concluye que los criterios de clasificación son dependientes.

**4.4.9 Pruebas entre las variables calificación de lenguaje y calificación de matemáticas**

Para el análisis de contingencia realizado con estas dos variables se definieron los siguiente criterios de clasificación, para la variable calificación de lenguaje los criterios son **C** que corresponde a las calificaciones desde 0 a 44 puntos y el criterio **D** que contiene a las calificaciones entre 45 y 83 puntos.

#### **Tabla CXL**

####  **Criterios de clasificación para las variables calificación de lenguaje y calificación de matemáticas**

|  |  |  |
| --- | --- | --- |
|  | **Criterios de clasificación para la variable calificación de lenguaje** |  |
| **Criterios de clasificación para la variable calificación de matemáticas** | C | **D** | **Total** |
| **A** | 6654.97 | 3647.029 | **102** |
| **B** | 2435.029 | 4129.97 | **75** |
| **Total** | **90** | **77** | **167** |

Los criterios de la variable calificación de matemáticas son **A** que corresponde a las calificaciones desde 0 a 14 puntos, el criterio **B** que contiene a las calificaciones entre 15 y 32 puntos. Las observaciones que cumplen con estos criterios se muestran en la tabla CXL.

Para determinar la independencia de los criterios de clasificación de las variables calificación de lenguaje y calificación de matemáticas se realizó el siguiente contraste de hipótesis:

**H0**: Los criterios de clasificación de las variables calificación de

lenguaje y calificación de matemáticas son independientes

vs.

**Ha**: Los criterios de clasificación de las variables calificación de

lenguaje y calificación de matemáticas son dependientes

Valor del estadístico de prueba χ2= 12.332

Grados de libertad 1

Valor p = 0.0004455

Se concluye que existe evidencia estadística para rechazar H0, en favor de Ha por lo tanto los criterios de clasificación son dependientes.

### 4.4.10 Resultados del análisis de contingencia entre otras variables aleatorias

Los resultados obtenidos al realizar el análisis de contingencia entre otras variables aleatorias, se muestran en la tabla CXLI, en la cual se indica entre que variables aleatorias se efectuó el análisis de contingencia, el valor p de la prueba y la conclusión.

**Tabla CXLI**

**Resultados del análisis de contingencia realizado a otras variables aleatorias**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable X** | **Variable Y** | **Valor p** | **Conclusión** |
| Suma de enteros | Resta de enteros | 0.004 | Dependientes |
| División de fracciones | Multiplicación de fracciones | 0.000 | Dependientes |
| Área del circulo | Área del triangulo | 0.000 | Dependientes |
| Ecuación con una incógnita | Ecuación con dos incógnitas | 0.033 | Dependientes |
| Estadística | Probabilidad | 0.104 | Independientes |
| Resta de enteros | Lectura Comprensiva | 0.713 | Independientes |
| Suma de enteros | Lectura Comprensiva | 0.000 | Dependientes |
| Palabras agudas | Palabras graves | 0.000 | Dependientes |
| Palabras sinónimas | Palabras antónimas | 0.000 | Dependientes |
| Palabras diminutivas | Palabras despectivas | 0.000 | Dependientes |
| Identificación del sujeto | Identificación del predicado | 0.000 | Dependientes |
| Identificación del sustantivo | Identificación del verbo | 0.000 | Dependientes |
| Identificación de frases | Identificación de oraciones | 0.000 | Dependientes |
| Palabras Agudas | Palabras tildadas | 0.037 | Dependientes |

Como se puede observar en la tabla CXLI, de los pares de variables aleatorias analizados, se obtuvo que las variables correspondientes a los ejercicios de estadística y de probabilidad son independientes; y entre las variables correspondientes a la resta de números enteros y lectura comprensiva también se obtuvo que son independencia no lineal.