[**4.5 Análisis de varianza** 312](#_Toc513608479)

[4.5.1Análisis de varianza de la variable aleatoria calificación de lenguaje con respecto a la ubicación de los planteles educativos fiscales rurales del cantón Guayaquil. 320](#_Toc513608480)

[4.5.2 Análisis de varianza de la variable aleatoria calificación de lenguaje con respecto a la jornada de los planteles educativos fiscales rurales del cantón Guayaquil. 322](#_Toc513608482)

## **4.5 Análisis de varianza**

El diseño de experimentos es una técnica estadística en la cual una variable aleatoria cuantitativa es explicada en términos de una o más variables aleatorias cualitativas. A las variables aleatorias cualitativas se las denomina **factores**, y a los distintos valores que pueden tomar dichos factores se los denomina **niveles o tratamientos**, el número de tratamientos de un factor se denota por a. La respuesta que se observa en cada uno de los tratamientos es una variable aleatoria.

Los datos típicos para un modelo unifactorial se tabulan como se muestra en la siguiente tabla:

|  |  |  |  |
| --- | --- | --- | --- |
| ***Tratamiento******(nivel)*** | ***Observaciones*** | ***Total*** | ***Promedio*** |
|  1 |  Y11 |  Y12 | … |  Y1n |  Y1. |   |
|  2 |  Y21 |  Y22 | … |  Y2n |  Y2. |   |
| … | … | … |  | … | … | … |
|  a |  Ya1 |  Ya2 | … |  Yan |  Ya. |   |
|  |  Y.. |   |

Estas observaciones se pueden escribir por medio un modelo estadístico lineal como el siguiente:

donde yij es la j-ésima observación del i-ésimo tratamiento, es un parámetro común de todos los tratamientos denominado media global, i es un parámetro único para el i-ésimo tratamiento llamado efecto del tratamiento i-ésimo, y ij es la componente aleatoria del error, donde ij ~ N(0, σ2) y se supone es constante para todos los tratamientos.

Una vez planteado el modelo matemático el objetivo es construir contrastes de hipótesis con respecto a los efectos de los tratamientos y hacer una estimación de estos. A este modelo se lo denomina *análisis de varianza* de clasificación en un sentido, dado que es un modelo unifactorial. Este diseño es completamente aleatorizado, por lo que se requiere que el experimento se realice en orden aleatorio, de tal forma que el medio ambiente en el que se usan los tratamientos tengan las mismas condiciones.

En el modelo unifactorial Yij = + i + ij, cuando σ2, la varianza del error ij permanece constante y se incluyen todos los a tratamientos del factor, el modelo se denomina *de efectos fijos.*En este modelo los efectos de los tratamiento i se definen como desviaciones con respecto a la media general , por lo que:

Sea yi el total de las observaciones bajo el j-ésimo tratamiento,  el promedio de las observaciones bajo el i-ésimo tratamiento. De igual forma, sea y.. la suma de todas las observaciones y la media general de las observaciones, expresado matemáticamente:

En donde i = 1, 2, ..., a y N = an es el número total de observaciones. La notación del punto en el subíndice, indica la suma del subíndice que reemplaza.

La media del i-ésimo tratamiento es E(yij)= i = + i, i =1, 2,. .., a. Es decir que el valor medio del i-ésimo tratamiento es la suma de la media general y el efecto del i-ésimo tratamiento. El contraste de hipótesis es:

H0 : μ1 = μ2 = ... = μa

vs.

H1 : μi = μj , para al menos un para de ( i, j)

El contraste se lo puede escribir en términos de los efectos de los tratamientos de la siguiente forma

H0 : τ1 = τ2 = ... = τa

vs.

H1 : τi ≠ 0, para al menos un i

Por lo tanto se desea probar la igualdad de las medias de los tratamientos, o que los efectos de los tratamientos τi son iguales a cero. El procedimiento que se utiliza para esta prueba es el análisis de varianza, es decir la prueba se basa en un análisis de la variabilidad total de los datos, la cual está dada por la *suma total de cuadrados* *SCT*, si se divide SCT para an-1 se obtiene la varianza de la población, la cual es una medida de la variabilidad de las observaciones.

SCT =  , donde 

La suma total de cuadrados se puede expresar como la *suma de cuadrados de los tratamientos* *SCTr* y la *suma de cuadrados del error SCE*, como se muestra a continuación,



La suma cuadrática de los tratamientos SCTr, mide la variación existente entre las medias de los a tratamientos, mientras que la suma cuadrática del error SCE, mide la variación existente dentro de cada tratamiento.

Dado que los Yi1, Yi2, ..., Yin , para i = 1, 2, ... , a representan una muestra aleatoria de tamaño n, tomada de una población normal con media + i y varianza σ2 , se tiene que para cada i



es una variable aleatoria ji-cuadrado χ2 con n-1 grados de libertad. Además puesto que los a tratamientos son variables aleatorias independientes, se tiene que



es una variable aleatoria ji-cuadrado con a(n-1) grados de libertad.

Luego supongamos que X es una variable aleatoria ji-cuadrado con ν grados de libertad, entonces E (X) = ν, en base a esto el valor esperado de la variable aleatoria (1/σ2) SCE es a(n-1), donde  es un estimador de σ2, esta cantidad se denomina *cuadrado medio del error* y se denota por *CME*.

Por otro lado se tiene que las  son variables aleatorias independientes, distribuidas normalmente con media μ y varianza , se tiene que



es una variable aleatoria ji-cuadrado con a-1 grados de libertad. El valor esperado de esta variable aleatoria es igual a (a-1), por lo tanto es otro estimador de σ2, este estimador se denomina *cuadrado medio de los tratamientos*y se denota por *CMTr*.

En este análisis se supone que los estimadores de σ2 , CMTr y CME son variables aleatorias independientes con a-1 y a(n-1) grados de libertad respectivamente, entonces  es una variable aleatoria que tiene distribución F con (a-1) y a(n-1) , grados de libertad.

La hipótesis nula H0 (de que los efectos de los a tratamientos son iguales a cero) se rechaza si el valor *f* calculado en base a las observaciones, es mayor que Fα, k-1, k(n-1), donde α es el nivel de significancia de la prueba. Si la hipótesis nula es verdadera CMTr es un estimador insesgado de σ2.

**Tabla CXLII**

**Análisis de varianza de clasificación en un sentido**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fuente** | **Grados de libertad** | **Sumas cuadráticas** | **Cuadrado medio** | **F** |
| Tratamientos | a-1 | SCTr | CMTr |  |
| Error | a(n-1) | SCE | CME |  |
| Total | an-1 | STC |  |  |

El procedimiento descrito anteriormente se denomina análisis de varianza de clasificación en un sentido y se lo representa mediante una *tabla de análisis de varianza ANOVA,* como se muestra en la tabla CXLII. Los diferentes métodos de hacer comparaciones múltiples se emplean sólo cuando el resultado del ANOVA resulta significativo. En tal caso, se sabe que existen diferencias entre las muestras, pero sin poder especificar entre cuales de ellas. Se necesita, entonces, alguna forma de poder compararlas entre sí, y alcanzar así el objetivo final del ANOVA. Los métodos de comparaciones múltiples son la herramienta adecuada para este fin, pues permiten un análisis muy sutil de la significación encontrada. Para los casos en que se determine que los efectos de los tratamientos no son todos iguales a cero, se utilizará el método de la *mínima diferencia significativa (LSD)*, para determinar en que parejas de tratamientos existe diferencias de medias.

Con el método *LSD* se desea probar el siguiente contraste de hipótesis

H0 : μi - μj = 0 , para i ≠ j, e i, j = 1, 2, ... , a

vs.

H1 : μi - μj ≠ 0 , para i ≠ j, e i, j = 1, 2, ... , a

Paro lo cual se utiliza el estadístico de prueba:



Para probar esta hipótesis se calcula la mínima diferencia significativa LSD, tal como sigue,

, donde N = a x n, entonces se rechaza la hipótesis nula H0 en favor de la hipótesis alterna Ha sí, .

### 4.5.1Análisis de varianza de la variable aleatoria calificación de lenguaje con respecto a la ubicación de los planteles educativos fiscales rurales del cantón Guayaquil.

El número de tratamientos del factor ubicación es cinco, los cuales corresponden a las parroquias rurales Progreso, Sabana Grande, Puná, Posorja y Tenguel. Para realizar el análisis de varianza entre la variable aleatoria calificación de lenguaje con respecto a la ubicación de los planteles educativos fiscales rurales del cantón Guayaquil, se planteó el modelo estadístico lineal siguiente:

donde yij es la j-ésima observación del i-ésimo tratamiento, es 43.14 puntos, i es el efecto del i-ésimo tratamiento y ij es la componente aleatoria del error, donde ij ~ N(0, σ2) y se supone, es constante para todos los tratamientos.

El contraste de hipótesis es:

H0 : μ1 = μ2 = μ3 = μ4 = μ5

vs.

H1 : μi = μj , para al menos un para de ( i, j)

El contraste en términos de los efectos de los tratamientos es

H0 : τ1 = τ2 = τ3 = τ4 = τ5

vs.

H1 : τi ≠ 0, para al menos un i

Por lo tanto se desea probar la igualdad de las medias de los tratamientos, o que los efectos de los tratamientos τi son iguales a cero. En el análisis de varianza de la variable aleatoria calificación de lenguaje con respecto al factor ubicación se obtuvo los siguientes resultados:

**Tabla de análisis de varianza**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Fuente*** | ***Grados de libertad*** | ***Sumas cuadráticas*** | ***Cuadrados medios*** | ***F*** | ***P*** |
| ***Tratamientos*** | *4* | *355.86* | *88.97* | *0.408* | *0.803* |
| ***Error*** | *162* | *35340.655* | *218.152* |  |  |
| ***Total*** | *166* | *35696.515* |  |  |  |

El valor p = 0.803 de esta prueba indica que existe evidencia estadística para no rechazar la hipótesis nula de que las medias de los a tratamientos son iguales. En el gráfico 4.7 se puede apreciar las medias de cada uno de los tratamientos del factor ubicación, con respecto a la calificación de la prueba de lenguaje.

**Gráfico 4.7**

***Medias de los tratamientos del factor ubicación con respecto a al calificación de lenguaje***

### 4.5.2 Análisis de varianza de la variable aleatoria calificación de lenguaje con respecto a la jornada de los planteles educativos fiscales rurales del cantón Guayaquil.

El número de tratamientos del factor jornada es tres, los cuales corresponden a las jornadas matutina, vespertina y nocturna. Para realizar el análisis de varianza entre la variable aleatoria calificación de lenguaje con respecto a la jornada de trabajo de los planteles educativos fiscales rurales del cantón Guayaquil, se planteó el modelo estadístico lineal siguiente:

donde yij es la j-ésima observación del i-ésimo tratamiento, es 43.14 puntos, i es el efecto del i-ésimo tratamiento y ij es la componente aleatoria del error, donde ij ~ N(0, σ2) y se supone, es constante para todos los tratamientos.

El contraste de hipótesis es:

H0 : μ1 = μ2 = μ3

vs.

H1 : μi = μj , para al menos un para de ( i, j)

El contraste en términos de los efectos de los tratamientos es

H0 : τ1 = τ2 = τ3

vs.

H1 : τi ≠ 0, para al menos un i

Por lo tanto se desea probar la igualdad de las medias de las jornadas matutina, vespertina y nocturna, o que los efectos de los tratamientos τi son iguales a cero. En el análisis de varianza de la variable aleatoria calificación de lenguaje con respecto al factor jornada se obtuvo los siguientes resultados:

**Tabla de análisis de varianza**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Fuente*** | ***Grados de libertad*** | ***Sumas******cuadráticas*** | ***Cuadrados medios*** | ***F*** | ***P*** |
| ***Tratamientos*** | *2* | *27.405* | *13.703* | *0.063* | *0.939* |
| ***Error*** | *164* | *35669.146* | *217.495* |  |  |
| ***Total*** | *166* | *35796.551* |  |  |  |

El valor p = 0.803 de esta prueba indica que existe evidencia estadística para no rechazar la hipótesis nula de que las medias de los a tratamientos son iguales.

**Gráfico 4.8**

***Medias de los tratamientos del factor jornada con respecto a la calificación de lenguaje***

En el gráfico 4.8 se puede apreciar las medias de cada uno de los tratamientos del factor jornada, con respecto a la calificación de la prueba de lenguaje. En el gráfico se puede observar que los planteles educativos que funcionan en jornada vespertina, obtuvieron la más alta media en la calificación de lenguaje.