## 4.6 Análisis de varianza

El Análisis de varianza es una técnica estadística que trata de explicar una variable cuantitativa en términos de una o más variables cualitativas, estas variables cualitativas son llamadas "**factores**" y cada uno de estos factores puede tener a niveles llamados "**tratamientos**", la respuesta que se observa en cada uno de los a tratamientos es una variable aleatoria. Por medio de esta técnica se formularán pruebas de hipótesis acerca de los efectos de los tratamientos, y se hará una estimación de ellos. El modelo a usar es el siguiente:

∈i  N(0,2) cov(∈i,∈j)=0 para ij

En donde yij es la j-ésima observación sometida al i-ésimo tratamiento del factor, 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. Para probar la hipótesis, se supone que los errores del modelo son variables aleatorias independientes con distribución normal, con media cero y varianza constante.

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, por lo tanto este diseño experimental es un diseño completamente aleatorizado.

En el modelo que se expuso anteriormente, se incluyen todos los a niveles del factor, por lo tanto es un *modelo de efectos fijos*.

En este modelo los efectos de tratamiento i se definen como desviaciones con respecto a la media general, por esta razón

Sea yi. el total de las observaciones bajo el j-ésimo tratamiento, yi. el promedio de las observaciones bajo el i-ésimo tratamiento. Similarmente, sea y.. la suma de todas las observaciones y 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. La prueba de hipótesis que se plantea (para probar la igualdad de los a tratamientos) es

Para al menos un i

Ahora con todo lo previo definido se procede a elaborar la tabla ANOVA.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fuente de Variación | Suma de cuadrados | Grados de libertad | Media de cuadrados | F0 |
| Tratamientos | SCTratamientos | a-1 | MCTratamientos | MCTratamientos/MCE |
| Error | SCE | N-a | MCE |  |
| Total | SCT | N-1 |  |  |

Donde:

y se rechaza la hipótesis nula con (1-α)100% si F0 > F, a-1, N-a.

Una vez que se ha rechazado la hipótesis nula (si se rechaza), se desea comparar todas las parejas de "a" medias de tratamientos, un método utilizado para este fin es el Método de la Mínima Diferencia Significativa (LSD). El contraste de hipótesis que se hace es el siguiente: H0: i = j, para toda i≠j. Esto se puede hacer empleando el estadístico t.

La pareja de medias i, j se consideran diferentes si:

La cantidad

Se denomina mínima diferencia significativa.

Para el caso del presente estudio se utilizará un modelo factorial de dos factores con interacciones, la tabla ANOVA cambia en forma, pero los resultados igual se interpretan como en el caso anterior, la única diferencia es que ahora se hace extensivo a dos factores. Se supone que los errores del modelo son variables aleatorias independientes con distribución normal, con media cero y varianza constante.

El modelo a usar es el siguiente:

∈ijk  N(0,2) cov(∈i,∈j)=0 para ij

Y la tabla de contingencia ahora es:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fuente de Variación | Suma de cuadrados | Grados de libertad | Media de cuadrados | F0 |
| Tratamiento A | SCA | a-1 | MCA | MCA/MCE |
| Tratamiento B | SCB | b-1 | MCB | MCB/MCE |
| Interacción | SCAB | (a-1)(b-1) | MCAB | MCAB/ MCE |
| Error | SCE | ab(n-1) | MCE |  |
| Total | SCT | abn-1 |  |  |

Donde:

Los contrastes de hipótesis ahora se hacen sobre los dos factores y la interacción.

Para al menos un i

Para al menos un i

(τβ)ij ≠ 0 para al menos un par de ij

**4.6.1 Análisis de varianza del modelo bifactorial para la variable nota de matemáticas**

Para el modelo bifactorial de la nota de matemáticas, el primer factor i es la jornada que tiene tres niveles: 1 (matutino), 2 (vespertino), 3 (nocturno) y el segundo factor βj es el sexo el cual tiene dos factores: 1 (femenino), 2 (masculino) y las interacciones son entre la jornada y el sexo, es decir que habrán seis interacciones.

∈ijk  N(0,2) cov(∈i,∈j)=0 para ij

La tabla ANOVA obtenida para este modelo se presenta en la tabla XCI

**Tabla XCI**

**Tabla ANOVA para el modelo de la nota de matemáticas explicado por los factores jornada y sexo**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fuente de Variación** | **Sumas cuadráticas** | **Grados de libertad** | **Medias cuadráticas** | **F** | **Valor p** |
| JORNADA | 3331.923 | 2 | 16665.961 | 18.327 | .000 |   |
| SEXO | 2.743 | 1 | 2.743 | .030 | .862 |   |
| JORNADA \* SEXO | 2840.069 | 2 | 1420.034 | 15.622 | .000 |   |
| Error | 99809.154 | 1098 | 90.901 |   |   |   |
| Total | 108702.533 | 1103 |   |   |   |   |

Con los valores p obtenidos se puede concluir que son significativos para la nota obtenida en matemáticas el factor jornada y la interacción de los factores sección y sexo. Aunque cabe recalcar que no es un muy buen modelo debido a que la media cuadrática del error es alta.

Para saber si existen diferencias entre los niveles del factor sección se realizó la prueba LSD y los resultados obtenidos se presentan en la tabla XCII

**Tabla XCIII**

**Prueba LSD para los niveles del factor jornada en el modelo para la nota de matemáticas**

|  |  |  |  |
| --- | --- | --- | --- |
|  **(I) tipo de jornada del colegio** | **(J) tipo de jornada del colegio** | **Diferencia de medias (I-J)** | **Valor p** |
| MAT | VES | 2.4000\* | .000 |   |
|  | NOC | 7.4906\* | .000 |   |
| VES | MAT | -2.4000\* | .000 |   |
|  | NOC | 5.0906\* | .000 |   |
| NOC | MAT | -7.4906\* | .000 |   |
|  | VES | -5.0906\* | .000 |   |

\* La diferencia de medias es significativa al nivel .05

Los valores p obtenidos nos indican que se rechazan la hipótesis nula de los contrastes de hipótesis formulados acerca de la igualdad de las medias de las notas obtenidas en matemáticas en las jornadas (de dos en dos), lo cual nos indica que existen diferencias en cuanto a las jornadas en la nota de matemáticas, para poderlo visualizar mejor se puede apreciar el gráfico 4.1.

**Gráfico 4.1**

**Diferencia entre las jornadas en la Nota de matemáticas**

Se puede observar en el gráfico que los alumnos de la jornada matutina obtienen las mejores calificaciones que los de las jornadas vespertinas y nocturnas, existe una gran diferencia en cuanto a las notas obtenidas en matemáticas, de acuerdo a la jornada.

En los resultados de la tabla ANOVA para la nota de matemáticas se aprecia que el sexo como factor no es influyente, pero si lo es su interacción con la jornada, mientras que en las tablas de contingencia se apreció que el sexo es un factor influyente en la calificación de matemáticas, esto se debe a que el análisis propuesto en esta sección es en base a un modelo lineal, mientras que las tablas de contingencia nos permiten saber si existe la relación de otro tipo (no lineal). Como resultados de la sección 4.3 se da a notar que el sexo femenino obtiene las mejores calificaciones en matemáticas, esto puede ser apreciado en el gráfico 4.2.

**Gráfico 4.2**

**Diferencia entre los sexos en la Nota de matemáticas**

En realidad en el gráfico se aprecia que existe diferencia en cuanto a la nota de matemáticas (influida por el sexo), se puede observar que el sexo femenino obtiene mejores calificaciones en matemáticas que el sexo masculino.

**4.6.2 Análisis de varianza del modelo bifactorial para la variable nota de lenguaje**

Para el modelo bifactorial de la nota de lenguaje, el primer factor i es la jornada que tiene tres niveles: 1 (matutino), 2 (vespertino), 3 (nocturno) y el segundo factor βj es el sexo el cual tiene dos factores: 1 (femenino), 2 (masculino) y las interacciones son entre la jornada y el sexo, es decir que habrán seis interacciones.

∈ijk  N(0,2) cov(∈i,∈j)=0 para ij

La tabla ANOVA obtenida para este modelo se presenta en la tabla XCIV

**Tabla XCIV**

**Tabla ANOVA para el modelo de la nota de lenguaje explicado por los factores jornada y sexo**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fuente de variación** | **Suma cuadráticas** | **Grados de libertad** | **Medias cuadráticas** | **F** | **Valor p** |
| JORNADA | 2857.066 | 2 | 1428.533 | 9.669 | .000 |   |
| SEXO | 1311.066 | 1 | 1311.066 | 8.874 | .003 |   |
| JORNADA \* SEXO | 2799.729 | 2 | 1399.865 | 9.475 | .000 |   |
| Error | 162227.634 | 1098 | 147.748 |   |   |   |
| Total | 176097.996 | 1103 |   |   |   |   |

Con los valores p obtenidos se puede concluir que son significativos para la nota obtenida en lenguaje el factor jornada, el factor sexo y la interacción de los factores jornada y sexo. Aunque cabe recalcar que no es un muy buen modelo debido a que la media cuadrática del error es alta.

Para saber si existen diferencias entre los niveles del factor sección se realizó la prueba LSD y los resultados obtenidos se los presenta en la tabla XCV

**Tabla XCV**

**Prueba LSD para los niveles del factor jornada en el modelo para la nota de lenguaje**

|  |  |  |  |
| --- | --- | --- | --- |
|  **(I) tipo de jornada del colegio** | **(J) tipo de jornada del colegio** | **Diferencia de medias (I-J)** | **Valor p** |
| MAT | VES | -4.9198\* | .000 |   |
|  | NOC | -2.6017\* | .029 |   |
| VES | MAT | 4.9198\* | .000 |   |
|  | NOC | 2.3181\* | .044 |   |
| NOC | MAT | 2.6017\* | .029 |   |
|  | VES | -2.3181\* | .044 |   |

\*La diferencia de medias es significativa al nivel .05

Los valores p obtenidos nos indica que se debe rechazar la hipótesis nula de los contraste formulados acerca de la igualdad de las media de notas obtenidas en lenguaje en las diferentes jornadas (de dos en dos), lo cual nos indica que existen diferencias en cuanto a las jornadas en la nota de lenguaje, para poderlo visualizar mejor se puede apreciar el gráfico 4.3.

**Gráfico 4.3**

**Diferencia entre las jornadas en la Nota de lenguaje**

Se aprecia en el gráfico 4.3 que los alumnos del décimo año de educación básica obtienen mejores calificaciones en lenguaje durante la tarde (jornada vespertina).

En los resultados de la tabla ANOVA para la nota de lenguaje se aprecia que el sexo si es un factor influyente, esto se puede apreciar en al gráfico 4.4

**Gráfico 4.4**

**Diferencia entre los sexos en la Nota de lenguaje**

Se puede apreciar que en cuanto a lenguaje el sexo femenino obtiene mejores calificaciones en lenguaje que el sexo masculino.

**4.6.3 Análisis de varianza del modelo bifactorial para la variable promedio general**

Para el modelo bifactorial de la variable promedio general, el primer factor i es la jornada que tiene tres niveles: 1 (matutino), 2 (vespertino), 3 (nocturno) y el segundo factor βj es el sexo el cual tiene dos factores: 1 (femenino), 2 (masculino) y las interacciones son entre la jornada y el sexo, es decir que habrán seis interacciones. El modelo a usar se presenta a continuación:

∈ijk  N(0,2) cov(∈i,∈j)=0 para ij

La tabla ANOVA obtenida para este modelo se presenta en la tabla XCVI

**Tabla XCVI**

**Tabla ANOVA para el modelo del promedio general explicado por los factores sexo y jornada**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fuente de Variación** | **Sumas cuadráticas** | **Grados de libertad** | **Medias cuadráticas** | **F** | **Valor p** |
| JORNADA | 125.530 | 2 | 62.765 | .926 | .396 |   |
| SEXO | 309.092 | 1 | 309.092 | 4.562 | .033 |   |
| JORNADA \* SEXO | 2840.396 | 2 | 1420.198 | 20.962 | .000 |   |
| Error | 74389.277 | 1098 | 67.750 |   |   |   |
| Total | 80210.008 | 1103 |   |   |   |   |

Con los valores p obtenidos se puede concluir que el factor sexo es influyente para el promedio y también la interacción entre sexo y jornada, se puede apreciar que la jornada no es significativa, por lo tanto se realizará un modelo unifactorial con la jornada como factor de explicación para determinar si en realidad la jornada no es un factor significativo.

**Tabla XCVII**

**Tabla ANOVA para el modelo del promedio general explicada solo con el factor sexo**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fuente de Variación** | **Sumas cuadráticas** | **Grados de libertad** | **Medias cuadráticas** | **F** | **Valor p** |
| JORNADA | 1499.192 | 2 | 749.596 | 10.455 | .000 |   |
| Error | 79078.825 | 1103 | 71.694 |   |   |   |
| Total | 80578.018 | 1105 |   |   |   |   |

 Se puede apreciar en la tabla anterior que la jornada sí es un factor influyente en el promedio de las notas, por este motivo se realizaron pruebas LSD para saber si hay diferencia entre los niveles y los resultados son los muestra en la tabla XCVIII

**Tabla XCVIII**

**Prueba LSD para los niveles del factor jornada en el modelo para promedio general**

|  |  |  |  |
| --- | --- | --- | --- |
| **(I) tipo de jornada del colegio** | **(J) tipo de jornada del colegio** | **Diferencia de medias (I-J)** | **Valor p** |
| MAT | VES | -1.2035\* | .016 |   |
|  | NOC | 2.4445\* | .003 |   |
| VES | MAT | 1.2035\* | .016 |   |
|  | NOC | 3.6479\* | .000 |   |
| NOC | MAT | -2.4445\* | .003 |   |
|  | VES | -3.6479\* | .000 |   |

\* La diferencia de medias es significativa al nivel .05

Se aprecia en la tabla que hay diferencias significativas en los niveles del factor jornada para el modelo del promedio general, a continuación se presentan dos gráficos para poder visualizar la diferencia existente entre los niveles de los factores.

**Gráfico 4.5**

**Diferencia entre las secciones en el promedio general**

**Gráfico 4.6**

**Diferencia entre los sexos en el promedio general**

En los gráficos anteriores se aprecia que la jornada vespertina sin ser la que mejor notas obtuviere en matemáticas y siendo la de mejores notas en lenguaje, es la que mejor promedio tiene, esto puede ser confirmado con los modelos que se presentaron con anterioridad. En cuanto a sexos una vez más se observa que el sexo femenino tiene mejor calificación (en este caso promedio).