ANÁLISIS DE VARIANZA

Dr. Cs. Ruiz Aranibar, Gustavo

gustavoruiz432@hotmail.com

RESUMEN

Los diferentes modelos del análisis de varianza, muestran cómo se desarrollan y aplican en la vida real, los cuales servirán para realizar investigaciones, analizar e interpretar los resultados. Se realiza un análisis del diseño experimental: el completamente aleatorizado y el de bloques completos y aleatorizados, se introduce el concepto de mediciones repetidas y el experimento factorial; como se utiliza con el diseño completamente aleatorizado, además se ilustra con ejemplos reales al análisis de datos aplicados que son útiles para la comprensión de los diferentes modelos, dado que el análisis de varianza es una de las técnicas estadísticas más utilizadas, con aplicaciones en todos los ámbitos del saber humano.

PALABRAS CLAVE

Contraste, diseño factorial, docimasia, factores, homocedasticidad, interacción, niveles, tratamiento.

1. INTRODUCCIÓN

Las Técnicas Iniciales del Análisis de Varianza (ANAVA) fueron desarrolladas originalmente por el estadístico y genetista Sir Ronald A. Fisher en los años 1920 y 1930, pionero e innovador del usos y aplicaciones de métodos estadísticos en el diseño experimental, usando la distribución F de Fisher como parte del contraste de hipótesis, cuyas aplicaciones hoy en día se realizan en todas las disciplinas científicas, permitiendo la realización de pruebas de hipótesis de significación para determinar qué factores influyen en el resultado del experimento, donde la comprobación de hipótesis es una característica muy útil del ANAVA.

En 1919 Fisher empezó a trabajar en la Rothamsted Experimental Station (Harpenden, Hertfordshire, Inglaterra). allí comenzó el estudio de una extensa colección de datos, cuyos resultados fueron publicados bajo el título general de Studies in Crop Variation. Durante los siguientes siete años, se dedico al estudio pionero de los principios del Diseño de Experimentos, elaboro sus trabajos sobre el ANAVA (Harpenden, Hertfordshire, Inglaterra), y comenzó a prestar una atención especial a las ventajas metodológicas de la computación de datos (Statistical Methods for Research Workers, 1925).

En estadística, el ANAVA es una colección de modelos estadísticos y sus procedimientos asociados, en el cual la varianza esta particionada en ciertos componentes debido a diferentes variables explicativas. Un ANAVA es ejecutado para un diseño factorial mediante el uso de operadores especiales sugeridos por H. O. Hartley.

El ANAVA compara diferencias en los resultados entre varios grupos y en muchas aplicaciones industriales, que implican experimentos en los que se toman en cuenta solamente los grupos o niveles pertenecientes a un factor de interés (temperatura, sabor) que puede tener varios niveles numéricos o categóricos (100º, 200º, 350º, etc. o vainilla, chocolate, frambuesa, piña, etc.). Este también compara las medias de cierto número de grupos o poblaciones, siendo su objetivo analizar las diferencias entre las medias de los grupos, esta variación total se subdivide en variación entre grupos, la que se considera como error experimental y variación dentro de los grupos que se le atribuye a efectos de tratamiento.

2. SUPUESTOS PREVIOS

El ANAVA parte de algunos supuestos o hipótesis que deben cumplirse como ser: La variable dependiente debe medirse al menos a nivel de intérvalo, independencia de las observaciones, la distribución de los residuales debe ser normal y la homocedasticidad.

En caso que alguno de estos supuestos (normalidad, homogeneidad de varianzas o independencia de los términos de error de cada observación) no se cumplan, impactarán sobre la distribución del estadístico F y con ello el verdadero nivel de significancia de la prueba de hipótesis del ANAVA, afectando así la calidad de las conclusiones que finalmente se obtendrá. Por ello, resulta importante verificar que los supuestos del análisis se cumplen antes de elaborar conclusiones.

3. DISEÑOS FACTORIALES

Diseño factorial o arreglo factorial, es un diseño experimental cuyas técnicas permiten realizar una planificación de las pruebas del experimento, diseño en el que se investigan todas las posibles combinaciones de los niveles de los factores en cada réplica del experimento para la obtención de numerosas conclusiones, permitiendo un ahorro económico de tiempo y trabajo.

El desarrollo de la metodología del ANAVA tiene influencia sobre los tipos de investigación experimental que se lleva a cabo en muchos campos, siendo una técnica estadística aplicada en ciencias sociales, agricultura, biología, medicina, investigación farmacéutica, psicología, toxicología, ingenierías de: minas, metalurgia, química, etc.

Los diseños factoriales permiten al investigador planificar un trabajo para evaluar el efecto combinado de dos o más variables de forma simultánea en el resultado medido, obteniéndose también información en cuanto a la posible interacción entre los diversos factores. El ANAVA es un método que permite comparar varias medias en diversas situaciones muy ligadas a los diseños experimentales.

En muchos experimentos es frecuente considerar dos o más factores y estudiar el efecto conjunto que dichos factores producen sobre la variable respuesta. Para resolver esta situación se utiliza el diseño factorial. En estos diseños, los factores que intervienen tienen la misma importancia a priori y se supone por tanto, la posible presencia de interacción.

Entendiéndose por factor una característica cualitativa de las condiciones experimentales, éste tiene diferentes niveles que son los valores cuantitativos, un tratamiento es el conjunto de niveles de los factores, una interacción es el resultado de la acción combinada de dos o más factores, el diseño de un experimento es una secuencia completa de pasos efectuados de antemano que conduzcan a conclusiones válidas del problema.

Las etapas que conducen a las conclusiones que ofrece el diseño factorial en este análisis será el ANAVA, dócimas de significación y medida de influencia de los efectos. Aplicar un ANAVA al diseño experimental, es el de estudiar la influencia de los factores y sus interacciones sobre las respuestas de un experimento.

Las dócimas de hipótesis se realizan por medio de la distribución F de Fisher, los valores calculados de la razón de F calculada, mostrados en una tabla de ANAVA, son comparados con los valores de las tablas de F (Fisher-Snedecor), considerando los respectivos grados de libertad tanto del numerador y del denominador; en esta comparación si el $F_{cal} > F_{teo}$ entonces se dice que es significativo, de lo contrario no lo es, a un determinado nivel de significación α o a un nivel de confianza 1-α.

4. CLASIFICACIÓN DEL ANÁLISIS DE VARIANZA

Cada método del ANAVA está asociado a un modelo matemático específico. Los modelos se clasifican según el número de variables que han de ser probadas. Si es una variable, el modelo se denomina de clasificación simple, de un solo factor o de una sola vía. Si son dos variables se denomina de clasificación doble o de dos factores y si tiene n variables será de clasificación múltiple o de n variables (MANAVA). Los modelos son:

1). Modelo de clasificación simple o experimentos de un factor

- a) Completamente aleatorizado
- b) Aleatorizado por bloque

2) Modelo de clasificación doble o experimentos de dos factores

- a) Sin replicación
- b) Con replicación

3) Modelo de clasificación múltiple

4.1. ANAVA DE UN FACTOR

Diseño completamente aleatorizado.

Sea X una característica que se mide en k poblaciones (o tratamientos) diferentes, con medias: $\mu_1, \mu_2, \ldots\ldots\ldots, \mu_k$ y varianzas respectivas: σ_1^2, σ_2^2, ..., σ_k^2 .

Se supone que las k poblaciones son independientes, cada una de ellas tiene distribución normal y las k varianzas son iguales a la varianza común σ^2 .

$\mu = \frac{\sum_{i=1}^{k} \mu_i}{k}$ Media total de las k poblaciones

Las variables aleatorias X_{ij} $i=1,2,\ldots\ldots,k$ $j=1,2,\ldots\ldots\ldots,n_i$ que denotan a la j-ésima observación de la i-ésima muestra son independientes y tienen una distribución normal $N(\mu_i, \sigma^2)$.

El modelo de clasificación de un factor completamente aleatorizado, parte del arreglo tabular como muestra la tabla siguiente:

	Tratamientos						
	1	2	...	i	...	k	
	x_{11} x_{12} . x_{1n_1}	x_{21} x_{22} . x_{2n_2}	... ...	x_{i1} x_{i2} . x_{in_i}	... ...	x_{k1} x_{k2} . x_{kn_k}	
Total T_i	T_1	T_2	...	T_i	...	T_k	$T.$
n_i	n_1	n_2	...	n_i	...	n_k	n
Medias	$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_i$		$\bar{x}_k$	$\bar{x}$

T_i = Suma de datos de la muestra i.

$T.$ = Total de datos de las k muestras.

$n_1+n_2+\ldots\ldots+n_k = n$ = Total observado de las k muestras.

$\bar{x}_i$ = Media de la muestra i, (estimación insesgada de la media u)

$\bar{x}_{..}$ = Media total muestral, (estimación insesgada de la media u)

En este estudio del ANAVA se tendrá: un modelo, una hipótesis y un análisis, de la siguiente relación:

$$x_{ij} - \bar{x}_{..} = x_{ij} - \bar{x}_{i.} + \bar{x}_{i.} - \bar{x}_{..}$$

Se obtiene la identidad de la suma de cuadrados:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_{i.} - \bar{x}_{..})^2$$

$$SCT \quad = \quad SCE \quad + \quad SCC$$

SCT= Suma de cuadrados total

SCE= Suma de cuadrados de error (o dentro de los tratamientos)

SCC= Suma de cuadrados de las columnas (o entre los tratamientos)

Las determinaciones de SCT, SCC y SCE, se calculan utilizando las siguientes equivalencias:

$$SCT = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}^2 - C \qquad C = \frac{T_{..}^2}{n}$$

$$SCC = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{k}n_i(\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{k}\frac{T_{i.}^2}{n_i} - C$$

$$SCE = SCT - SCC$$

Región crítica. Dado un nivel de significación a, para los grados de libertad *k-1* y *n-k*, en la tabla de la distribución F se encuentra el valor crítico $F_t = F_{i-\alpha,k-1,n-k}$, siendo la región crítica de la prueba el intervalo$]F_t, +\infty[$

Región de decisión. Se plantean dos hipótesis: las hipótesis nula H_0 y la alternativa H_1, a un nivel de significación α (o un nivel de confianza 1-α), teniéndose que comprobar:

$H_0: u_1 = u_2 = \ldots\ldots\ldots = u_k$
H_1 *: no todas las medias son iguales*

se rechaza H_0 si $F_{cal} > F_t$, y se acepta H_1 en caso contrario no rechazar H_0.

Es práctico resumir las sumas de cuadrados, los grados de libertad, los cuadrados medios y la F calculada en una tabla denominada ANAVA, como la siguiente:

Tabla 1
ANAVA. De un factor completamente aleatorizado

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Razón F calculada
Tratamientos (columnas)	*SCC*	*k-1*	$CMC = \frac{SCC}{k-1}$	$F_{cal} = \frac{CMC}{CME}$
Error	*SCE*	*n-k*	$CMC = \frac{SCE}{n-k}$	
Total	*SCT*	*n-1*		

4.2. ANAVA DE UN FACTOR. ALEATORIZADO POR BLOQUES

En el ANAVA de un factor aleatorizado por bloques completos, todos los tratamientos son asignados aleatoriamente a los bloques. Para un diseño de bloques completamente aleatorizados, se utiliza tres tratamientos, sean estos: $T_1 T_2, T_3$ asignados al azar a cuatro bloques, como sigue:

Bloque 1	Bloque 2	Bloque 3	Bloque 4
T_1	T_1	T_1	T_1
T_2	T_2	T_2	T_2
T_3	T_3	T_3	T_3

Los datos se registran en un arreglo como se muestra en la tabla:

Tabla 2
Registro de datos

	Tratamientos		
Bloques	1	2	3
1	x_{11}	x_{21}	x_{31}
2	x_{12}	x_{22}	x_{32}
3	x_{13}	x_{23}	x_{33}
4	x_{14}	x_{24}	x_{34}

Los k tratamientos A_i , i=1,2,…,k se asignan al azar a r bloques B_j , j=1,2,…,r , cuyos resultados en forma general se resumen en el arreglo r*k, tal como se muestra en la tabla:

	Tratamientos							
Bloques	**A1**	**A2**	**..**	**Ai**	**..**	**Ak**	**Total del bloque**	**Media de bloques**
B_1	x_{11}	x_{21}	..	x_{i1}	..	x_{k1}	$T_{.1}$	$\bar{x}_{.1}$
B_2	x_{12}	x_{22}	..	x_{i2}	..	x_{k2}	$T_{.2}$	$\bar{x}_{.2}$
:	:	:				:	:	
B_j	x_{1j}	x_{2j}	..	x_{ij}	..	x_{kj}	$T_{.j}$	$\bar{x}_{.j}$
:	:	:					:	
B_r	x_{1r}	x_{2r}	..	x_{ir}	..	x_{kr}	$T_{.r}$	$\bar{x}_{.r}$
Total	T_1	T_2	..	T_i	..	T_k	$T_{..}$	
Medias	$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_i$		$\bar{x}_k$		$\bar{x}_{..}$

$T_{1.}$: Suma de datos de la muestra i-ésima columna (tratamiento).

$T_{.1}$: Suma de datos de la j-fila (bloque)

$T_{..}$: Suma de las r k observaciones

$\bar{x}_{i.}$: Media de los r datos observados en el tratamiento A_1

$\bar{x}_{.j}$: Media de los k datos observados en el bloque B_j

$\bar{x}_{..}$: Media de todas las r k observaciones

Las X_{ij} son las variables aleatorias independientes y que cada una tienen distribución normal con media μ_{ij} y varianza común σ^2.

Modelo. Sea μ_i. la media de las r medias poblacionales para el tratamiento i, la media de las k poblaciones para el bloque j y la media total μ se las r^k poblaciones, lo que se representa por:

$$\mu_{i\,.} = \frac{\sum_{j=1}^{r} \mu_{ij}}{r} \qquad \mu_{.\,j} = \frac{\sum_{i=1}^{k} \mu_{ij}}{k}$$

$$\mu = \frac{\sum_{i=1}^{k} \sum_{j=1}^{r} \mu_{ij}}{rk} \qquad \begin{array}{l} i = 1,2,......,k \\ j = 1,2,......,r \end{array}$$

Cada observación X_{ij} puede escribirse de la forma:

$$X_{ij} = \mu_{ij} + \varepsilon_{ij}$$

En donde ε_{ij} mide la desviación del dato observado X_{ij} de la media poblacional μ_{ij}.

Como cada X_{ij} se supone normal $N(\mu_{ij},\sigma^2)$, entonces, las variables aleatorias ε_{ij} se suponen independientes y normales $N(0,\sigma^2)$.

Las desviaciones de μ_{ij} con respecto a μ se deben tanto a efectos de tratamiento como de bloques. Se supone que se suman los efectos, entonces:

$$X_{ij} = \mu_{ij} + \alpha_i + \beta_j$$

Donde α_i es el efecto del i-ésimo tratamiento y β_j es el efecto del j-iésimo bloque. Se impone la siguiente restricción:

$$\sum_{i=1}^{k} \alpha_i = 0 \quad \text{y} \quad \sum_{i=1}^{r} \beta_i = 0$$

Resumiendo, el modelo del ANAVA de un factor aleatorizado por bloques es la ecuación:

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Y donde se supone que:

1). Los efectos de los tratamientos y de los bloques son aditivos, es decir no hay efecto conjunto entre α_i y β_j . No hay interacción entre tratamientos y bloques.

2) Las variables aleatorias εij son independientes y normales $N(0,\sigma^2)$.

En el ANAVA de un factor aleatorizado por bloques, los bloques no deben ser considerados como un factor, ya que éstos son considerados solo como material experimental cuyos efectos influyen en los efectos de los tratamientos.

Hipótesis. Para determinar si hay diferencias significativas entre las medias poblacionales de los k tratamientos (columnas) la hipótesis nula y la alternativa respectivamente son:

H_0^c : $\mu_1 = \mu_2 = ... = \mu_{k.}$

H_1^c : *no todas las* $\mu_{i..}$ *son iguales*

Para determinar si hay diferencias significativas entre las medias poblacionales de los r bloques (filas) las hipótesis: nula y alternativa respectivamente son:

H_0^F : $\mu_1 = \mu_2 = \ldots\ldots = \mu_{.k}$

H_1^F : *no todas las* $\mu_{.j}$ *son iguales*

Por otra parte, si se supone que $\sum_{i=1}^{k} \alpha_i = 0$ y $\sum_{i=1}^{r} \beta_i = 0$ entonces:

$$\mu_{i..} = \frac{\sum_{j=1}^{r} \mu_{ij}}{r} = \frac{\sum_{j=1}^{r} (\mu + \alpha_i + \beta_j)}{r} = \mu + \alpha_i$$

$$\mu_{.j} = \frac{\sum_{i=1}^{k} \mu_{ij}}{k} = \frac{\sum_{i=1}^{k} (\mu + \alpha_i + \beta_j)}{k} = \mu + \beta_j$$

Las hipótesis respectivas en función de los efectos de tratamientos y de los bloques son:

Para tratamientos (columnas)

H_0^c : $\alpha_1 = \alpha_2 = \ldots\ldots = \alpha_{.k} = 0$

H_1^c : *al menos una de las* α_i *no es igual a cero*

Para los bloques (filas)

H_0^F : $\beta_1 = \beta_2 = \ldots\ldots = \beta_{k.} = 0$

H_1^F : *al menos una de los* β_i *no es igual a cero*

Análisis. De la identidad:

$$x_{ij} - \bar{x}_{..} = (x_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}) + (\bar{x}_{i.} - \bar{x}_{...}) + (\bar{x}_{.j} - \bar{x}_{..})$$

$$x_{ij} - \bar{x}_{..} = (x_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}) + (\bar{x}_{i.} - \bar{x}_{...}) + (\bar{x}_{.j} - \bar{x}_{..})$$

H_1^c : *al menos una de las* α_i *no es igual a cero*

Se obtiene:

$$\sum_{i=1}^{k}\sum_{J=1}^{r} (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{r} (x_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{r} (\bar{x}_{i.} - \bar{x}_{...})^2 + \sum_{i=1}^{k}\sum_{j=1}^{r} (\bar{x}_{.j} - \bar{x}_{..})^2$$

$$SCT = \sum_{i=1}^{k}\sum_{J=1}^{r} (x_{ij} - \bar{x}_{..})^2 \quad SCE = \sum_{i=1}^{k}\sum_{j=1}^{r} (x_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$$

$$SCC = \sum_{i=1}^{k}\sum_{j=1}^{r} (\bar{x}_{i.} - \bar{x}_{...})^2 \quad SCF = \sum_{i=1}^{k}\sum_{j=1}^{r} (\bar{x}_{.j} - \bar{x}_{..})^2$$

$$SCT = SCE + SCC + SCF$$

$$CMC = \frac{SCC}{k-1} \qquad CMF = \frac{SCC}{r-1} \qquad CME = \frac{SCE}{(k-1)(r-1)}$$

Para probar la hipótesis nula de que los efectos de los tratamientos son todos iguales a cero, se utiliza la estadística:

$F^C = \dfrac{CMC}{CME}$, que tiene distribución $F_{(k-1),\,(k-1)(r-1)}$

$F^F = \dfrac{CMC}{CME}$, que tiene distribución $F_{(r-1),\,(k-1)(r-1)}$

Se resume todas estas expresiones en el cuadro de ANAVA siguiente:

Tabla 4
ANAVA. Modelo de clasificación simple aleatorizado por bloques

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Razón F calculada
Tratamientos (columnas)	*SCC*	*k-1*	$CMC = \frac{SCC}{k-1}$	$F^c_{cal} = \frac{CMC}{CME}$
Entre bloques (filas)	*SCF*	*r -1*	$CMF = \frac{SCF}{r-1}$	$F^c_{cal} = \frac{CMF}{CME}$
Error	*SCE*	*(r-1)(k-1)*	$CME = \frac{SCE}{(r-1)(k-1)}$	
Total	*SCT*	*rk-1*		

Para calcular las sumas de cuadrados, se ejecutan las siguientes equivalencias:

$$SCT = \sum_{i=1}^{k}\sum_{J=1}^{r} (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{k}\sum_{J=1}^{r} x_{ij}^2 - C \qquad C = \frac{T_{..}^2}{rk}$$

$$SCC = \sum_{i=1}^{k}\sum_{j=1}^{r} (\bar{x}_{i..} - \bar{x}_{..})^2 = \frac{1}{r}\sum_{i=1}^{k} T_{i.}^2 - C$$

$$SCF = \sum_{i=1}^{k}\sum_{j=1}^{r} (\bar{x}_{.j.} - \bar{x}_{...})^2 = \frac{1}{k}\sum_{j=1}^{r} T_{.j}^2 - C$$

$$SCE = SCT - SCC - SCF$$

Los grados de libertad de SCE se obtienen también por sustracción:

$$(k - 1)(r - 1) = (rk - 1) - (k - 1) - (r - 1)$$

Nota. En el modelo simple aleatorio completamente aleatorizado por bloques solo interesa determinar si los efectos de los tratamientos son nulos o no. No es importante docimar si los efectos de los bloques son nulos o no, porque los bloques se consideran simplemente como material experimental. Sin embargo, esta segunda prueba es importante para el modelo de clasificación de dos variables sin repetición.

4.3. ANAVA DE DOS FACTORES

En este análisis se tiene un arreglo de k*r celdas, donde las columnas representan los niveles de un factor A y las filas los niveles de un factor B. El ANAVA de dos factores se clasifica según el número de observaciones en las celdas. Si cada celda contiene una sola observación de la muestra, el modelo se denomina sin réplica (o sin repetición) y si cada celda contiene dos o más observaciones de la muestra, el modelo se denomina con réplica (o con repetición).

4.4. ANAVA DE DOS FACTORES SIN REPLICACIÓN

Es similar al modelo de clasificación de un solo factor aleatorizado en bloques, además las dos variables son independientes, es decir que no hay interacción entre los dos factores.

4.5. ANAVA DE DOS FACTORES CON REPLICACIÓN

Es conocido como diseño completamente aleatorizado de dos factores, se caracteriza este modelo de clasificación de dos factores con replicas donde los tratamientos no son independientes y en este caso existe interacción de los dos factores, la interacción indica que los efectos de los niveles de un factor varia con los niveles del otro factor.

Este diseño se basa en un arreglo rectangular de las observaciones, en el que las c columnas representan los niveles o los tratamientos del factor A y las r filas los niveles o tratamientos del factor B. Cada combinación de tratamiento define una celda en la tabla, teniéndose r*c celdas, cada celda contiene n ($n \geq 2$) observaciones (replicas). Los r*c*n datos se muestran en la tabla siguiente:

Tabla 5
Experimento de dos factores con n réplicas

	Factor A				
Factor B	1	2	...	c	Total del bloque
1	x_{111} x_{112} . x_{11n}	x_{211} x_{212} . x_{21n}	... ...	x_{c11} x_{c12} . x_{c1n}	
	$T_{11.}$	$T_{21.}$	...	$T_{c1.}$	$T_{.1.}$
2	x_{121} x_{122} . x_{12n}	x_{221} x_{222} . x_{22n}	... ...	x_{c21} x_{c22} . x_{c2n}	
	$T_{12.}$	$T_{22.}$	...	$T_{c2.}$	$T_{.2.}$
....	....	....	...	....	....
R	x_{1r1} x_{1r2} . x_{1rn}	x_{2r1} x_{2r2} . x_{2rn}	... ...	x_{cr1} x_{cr2} . x_{crn}	
	$T_{1r.}$	$T_{2r.}$	...	$T_{cr.}$	$T_{.r.}$
	$T_{1..}$	$T_{2..}$	...	$T_{c..}$	$T_{...}$

Donde:

X_{ijk} es la k-ésima observación del i-ésimo nivel del factor A y de j-ésimo nivel del factor B (i=1,2,…,c, j=1,2,…,r k=1,2,…,n)

$T_{ij.}$: Suma de datos de la ij-ésima celda

$T_{i..}$: Suma de datos de la i-ésima columna

$T_{.j.}$: Suma de datos de la j-ésima fila

$T_{...}$: Suma de todas las rcn observaciones

$\bar{x}_{j..}$: Media de datos de la i-ésima columna

$\bar{x}_{.j.}$: Media de datos de la j-ésima fila

$\bar{x}_{...}$: Media de todas las r observaciones

Con el ANAVA de dos factores con réplicas, se prueban 3 hipótesis nulas distintas, estas son: que no existan efectos por columna (o las medias por columna no difieren significativamente), que no existan efectos por filas (o las medias por filas no difieran significativamente), y que no exista interacción entre los dos factores (los dos factores son independientes).

Modelo. El análisis de varianza de dos factores y con réplicas se basa en el modelo matemático:

$$X_{ijk.} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$
$$\forall\ i = 1,c \quad \forall\ j = 1,r \quad \forall\ k = 1,n$$

μ es la media global (sin importar el tratamiento)

α_i es el efecto del i-ésimo tratamiento del factor A

β_j es el efecto del j-ésimo tratamiento del factor B

γ_{ij} es el efecto de interacción del i-ésimo tratamiento del factor A y del j-ésimo tratamiento del factor B.

ε_{ijk} es el error aleatorio asociado al proceso de muestreo.

Se supone que: las variables aleatorias ε_{ijk} son independientes $N(0,\sigma^2)$ y que:

$$\sum_{i=1}^{c} \alpha_i = 0 \qquad \sum_{j=1}^{r} \beta_j = 0$$
$$\sum_{i=1}^{c} \gamma_{ij} = 0 \ \forall j \quad \sum_{j=1}^{r} \gamma_{ij} = 0 \ \forall i$$

Hipótesis. En el análisis de varianza de dos factores con replicación se realizan tres pruebas de hipótesis, estas son: para columnas, para filas y para interacción. Las hipótesis nulas y alternativas son las siguientes:

a). Los efectos de todos los tratamientos del factor A o de todas las columnas son nulos.

$$H_0^C : \alpha_i = 0 \quad \forall i = 1,c$$
$$H_1^C : \alpha_i \neq 0 \text{ para algunas columnas.}$$

b). Los efectos de todos los tratamientos del factor B o de todas las filas son nulos.

$$H_0^F : \beta_j = 0 \quad \forall j = 1,r$$
$$H_1^F : \beta_i \neq 0 \text{ para algunas filas.}$$

c). Los efectos conjuntos en todas las celdas son nulos, o no hay interacción entre filas y columnas.

$$H_0^I : \lambda_{ij} = 0 \ \ \forall i = 1,c \ \forall j = 1,r$$
$$H_1^I : \lambda_{ij} \neq 0 \text{ para algunas celdas.}$$

Análisis. Las estadísticas para realizar las tres pruebas de las hipótesis nulas dadas se obtienen de la partición de la suma de cuadrados siguientes:

$$\sum_{i=1}^{c}\sum_{j=1}^{r}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2 = rn\sum_{i=1}^{c}(\bar{x}_{i..} - \bar{x}_{...})^2$$
$$+ cn\sum_{j=1}^{r}(\bar{x}_{.j.} - \bar{x}_{...})^2$$
$$+ n\sum_{i=1}\sum_{j=1}(x_{ij.} - x_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2$$
$$+ \sum_{i=1}^{c}\sum_{j=1}^{n}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2$$

Esta partición se representa simbólicamente por:

$$SCT = SCC' + SCF + SCI + SCE$$

SCI es la suma de cuadrados debido a la interacción.

Los grados de libertad respectivos son:

$$rcn - 1 = (c - 1) + (r - 1) + (r - 1)(c - 1) + rc(n - 1)$$

Por otro lado, se verifica la distribución de las siguientes estadísticas:

$$F^C = \frac{SCC/(c-1)}{SCE/(rc(n-1))} = \frac{CMC}{CME} \approx F_{(c-1),\ rc(n-1)}$$

$$F^F = \frac{SCF/(r-1)}{SCE/(rc(n-1))} = \frac{CMF}{CME} \approx F_{(r-1),\ rc(n-1)}$$

$$F^I = \frac{SCI/(r-1)(c-1)}{SCE/(rc(n-1))} = \frac{CMI}{CME} \approx F_{(r-1)(c-1),\ rc(n-1)}$$

Estas estadísticas se utilizan para probar las hipótesis de columnas, de filas y de interacción respectivamente.

Tabla 6
ANAVA. Modelo de clasificación a dos factores con repetición

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Razón F calculada
Factor A (columnas)	SCC	$c\text{-}1$	$CMC = \frac{SCC}{c - 1}$	$F^C_{cal} = \frac{CMC}{CME}$
Factor B (filas)	SCF	$r\text{ -}1$	$CMF = \frac{SCF}{r - 1}$	$F^F_{cal} = \frac{CMF}{CME}$
Interacción A*B	SCI	$(r\text{-}1)(c\text{-}1)$	$CMI = \frac{SCI}{(r-1)(c-1)}$	
Error	SCE	rc(n-1)	$CME = \frac{SCE}{rc(n-1)}$	$F^I_{cal} = \frac{CMI}{CME}$
Total	SCT	$rk\text{-}1$		

Para calcular las sumas de cuadrados se ejecutan las siguientes equivalencias:

Se obtienen de la partición de la suma de cuadrados como siguen:

$$SCT = \sum_{i=1}^{c}\sum_{j=1}^{r}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\sum_{k=1}^{n}x_{ijk}^2 - C$$

$$C = \frac{T_{...}^2}{rcn}$$

$$SCC = rn\sum_{i=1}(\bar{x}_{i..} - \bar{x}_{...})^2 = \frac{1}{rn}\sum_{i=1}^{c}T_{i..}^2 - C$$

$$SCF = cn\sum_{j=1}^{r}(\bar{x}_{.j.} - \bar{x}_{...})^2 = \frac{1}{cn}\sum_{j=1}^{r}T_{.j.}^2 - C$$

$$SCE = \sum_{i=1}^{c}\sum_{j=1}^{r}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\sum_{k=1}^{n}x_{ijk}^2 - \frac{1}{n}\sum_{j=1}^{r}T_{ij.}^2$$

$$SCI = SCT - SCC - SCF - SCE$$

4.6. ANAVA DE N FACTORES CON REPLICACIÓN

Se hace notar que en el menú computacional denominado Anava de n factores, ejecuta los cálculos para el número de factores y niveles que se desee.

5. TRABAJO COMPUTACIONAL

Por todo lo anteriormente mencionado, se observa que en la determinación de los ANAVAS, son numerosos los cálculos a realizar, por esta razón, se facilita su determinación de acuerdo a las necesidades del usuario, por medio de la ejecución del siguiente menú, en el cual para cada uno de ellos se tiene un ejemplo numérico mostrado en el inciso 7, para mayor comprensión e interpretación de la teoría desarrollada.

UNIVERSIDADES UTO, UMSA, UAGRM ANÁLISIS DE VARIANZA Dr. Cs. Gustavo Ruiz Aranibar	
1	Anava de un factor. Muestras de tamaños iguales
2	Anava de un factor. Muestras de tamaños diferentes
3	Anava de un factor. Diseño aleatorizado por bloques
4	Anava de dos factores sin replicación
5	Anava de dos factores con replicación
6	Anava de tres factores con replicación
7	Anava de n factores
0	Salir del sistema

6. CONCLUSIONES

Se ha observado que a lo largo del estudio del ANAVA se requieren de demasiados cálculos, lo cual se evita con los programas computacionales desarrollados para cada uno de los temas señalados en el menú, facilitando enormemente este tedioso trabajo. La parte teórica y los ejemplos ayudan a los usuarios a comprender la utilización de los diferentes modelos los cuales permiten su aplicación y adaptación a trabajos reales de investigación o de la industria.

Todos los trabajos generan información que muchas veces no se las utiliza por no tener conocimiento de esta clase de modelos, teniendo en cuenta que se debe adaptar el modelo al problema y no el problema al modelo. De los modelos correspondientes aplicados a los datos se pueden obtener varios beneficios dentro el campo experimental como también beneficios económicos, para tomar decisiones estadísticas según los resultados obtenidos.

Se concluye que el ANAVA se aplica con la finalidad de analizar las diferencias o semejanzas significativas tanto de las medias como de las varianzas, donde una alta o baja razón F, implicarían la aceptación o rechazo de la hipótesis, y por otro lado, se revela el efecto que tiene una variable sobre la otra de acuerdo a su población en cuanto a su grado de predictibilidad.

7. APLICACIONES

El desarrollo de la metodología del ANAVA tiene influencia sobre los tipos de investigación experimental que se llevan a cabo en muchos campos, siendo una técnica estadística aplicada en ciencias sociales, agricultura, biología, medicina, investigación farmacéutica, psicología, toxicología, ingeniería de: minas, metalurgia, química, etc. Estos modelos de ANAVA se utilizan para abordar problemas técnicos y complejos en diferentes disciplinas.

A continuación, se muestran aplicaciones prácticas de acuerdo a cada uno de los tópicos mencionados en el menú y en cada uno de ellos, la fuente de datos es elaboración propia.

7.1. ANAVA DE UN FACTOR PARA MUESTRAS DE TAMAÑOS IGUALES.

Se desea comprar cinco máquinas de marcas diferentes para su uso en el ensamble de un producto particular, las que están siendo comparadas respecto a su velocidad. El experimento diseñado para determinar si hay diferencias en la velocidad promedio de cinco máquinas, se observan los tiempos empleados en producir seis artículos en forma aleatoria en cada máquina. Los tiempos registrados en segundos son conocidos. Determinar al nivel de significación del 5% o sí las máquinas llevan a cabo la tarea a la misma velocidad promedio.

Datos

	Máquinas				
Art.	1	2	3	4	5
1	19,20	18,70	12,50	20,30	19,90
2	18,70	14,30	14,30	22,50	24,30
3	21,30	20,20	8,70	17,60	17,60
4	16,50	17,60	11,40	18,40	20,20
5	17,30	19,30	9,50	15,90	18,40
6	22,40	16,10	16,50	19,00	19,10

Solución.

Anava de un factor para muestras de tamaños iguales

ANAVA				
Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados Medios	Prueba F
Máquinas	237.417	4	59.354	10.138
Error	146.371	25	5.855	
Total	**383.788**	**29**		

Decisión: F_{cal}=10,138 > $F_{5\%,4,25}$=2,76 por lo que se rechaza H_0 y se concluye que la hipótesis alternativa es verdadera. Es decir, que a un nivel del 5% hay diferencia significativa. Concluyendo, las máquinas no llevan a cabo la tarea a la misma velocidad promedio.

7.2. ANAVA DE UN FACTOR PARA MUESTRAS DE TAMAÑOS DIFERENTES.

En una universidad, en un semestre, cinco docentes enseñan a cinco paralelos una misma asignatura, cuyas calificaciones finales son registradas con calificaciones de 0 a 100, teniéndose 49, 35, 21, 37 y 40 estudiantes respectivamente. Al nivel de significación del 5%. ¿Se podrá concluir que existe una diferencia significativa en las calificaciones promedio obtenidas con los cinco docentes?

Datos.

	Docentes				
Nº Alumnos	1	2	3	4	5
1	64,00	80,00	78,00	68,00	58,00
2	66,00	30,00	32,00	69,00	52,00
3	57,00	30,00	60,00	68,00	10,00
4	21,00	30,00	10,00	20,00	5,00
5	59,00	30,00	73,00	20,00	52,00
6	63,00	40,00	51,00	54,00	5,00
7	26,00	40,00	75,00	63,00	52,00
8	65,00	52,00	67,00	55,00	5,00
9	68,00	30,00	79,00	20,00	5,00
10	59,00	52,00	16,00	51,00	55,00
11	41,00	35,00	31,00	51,00	70,00
12	70,00	45,00	71,00	41,00	53,00
13	68,00	40,00	54,00	51,00	60,00
14	74,00	30,00	72,00	56,00	52,00
15	62,00	30,00	30,00	51,00	55,00
16	69,00	45,00	60,00	65,00	52,00
17	42,00	40,00	64,00	51,00	74,00
18	53,00	51,00	19,00	44,00	40,00
19	75,00	35,00	14,00	60,00	20,00
20	62,00	45,00	67,00	20,00	62,00
21	53,00	30,00	16,00	57,00	20,00
22	65,00	51,00		20,00	66,00
23	61,00	45,00		54,00	30,00
24	64,00	60,00		51,00	20,00
25	58,00	51,00		54,00	52,00
26	18,00	40,00		20,00	70,00
27	49,00	40,00		84,00	52,00
28	70,00	30,00		75,00	61,00
29	74,00	51,00		30,00	70,00
30	53,00	60,00		55,00	54,00
31	51,00	55,00		20,00	40,00
32	67,00	30,00		51,00	70,00
33	65,00	51,00		20,00	36,00
34	62,00	30,00		51,00	80,00
35	65,00	35,00		52,00	70,00
36	64,00			60,00	55,00
37	65,00			51,00	31,00
38	64,00				52,00
39	66,00				77,00
40	62,00				30,00
41	57,00				
42	67,00				
43	67,00				
44	51,00				
45	66,00				
46	67,00				
47	68,00				
48	57,00				
49	57,00				

Solución.

Anava de un factor para muestras de tamaños diferentes.

ANAVA				
Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	Prueba F
Docentes	7.390,27	4	1.847,57	6,1044
Error	53.570,1	177	302,66	
Total	**60.960,36**	**181**		

Decisión: F_{cal}=6,1044 > F_{teo}=2,42 por lo que se rechaza H_0 , se concluye que la hipótesis alternativa es verdadera. Es decir, que a un nivel del 5% hay diferencia significativa. En conclusión, existe una diferencia significativa en las calificaciones promedio obtenidas por los cinco docentes.

7.3. ANAVA DE UN FACTOR PARA UN DISEÑO ALEATORIZADO POR BLOQUES.

La administración de una cadena de restaurantes de comida rápida tiene 5 sucursales en la ciudad, ésta desea evaluar el servicio de los restaurantes. Para lo cual se contrata a 35 investigadores (estimadores) con experiencia variada en evaluación del servicio de comida, siendo clasificados en siete bloques de cinco elementos. Los cinco estimadores son asignados de manera aleatoria para realizar la evaluación de servicio de un restaurante en particular, utilizando una escala que va de 0 (baja) a 100 (alta). Se desea probar las diferencias entre restaurantes a un nivel α = 5%.

Datos

Bloq.	Restaurantes				
Oper.	1,00	2,00	3,00	4,00	5,00
1	68,00	63,00	80,00	76,00	79,00
2	75,00	77,00	86,00	78,00	81,00
3	74,00	69,00	88,00	82,00	83,00
4	78,00	65,00	94,00	78,00	85,00
5	82,00	68,00	90,00	86,00	82,00
6	76,00	70,00	96,00	88,00	78,00
7	80,00	78,00	96,00	86,00	85,00

Solución

ANAVA de un factor para un diseño aleatorizado por bloques

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	Razón F calculada
Restaurantes	1.715,14	4	1.847,57	$F^C = 32,5595$
Evaluadores	303,94	6	302,66	$F^F = 3,8465$
Error	316,06	24	13,1693	
Total	**2.335,14**	**34**		

Decisión: $F_{cal}=32,5595 > F_{5\%,4,24}=2,78$ por lo que se rechaza H_0, se concluye que existe evidencia de una diferencia en el servicio de atención promedio entre los diferentes restaurantes.

Una verificación de la efectividad de la conformación de bloques, prueba una diferencia entre los grupos de estimadores, para lo cual se tiene:

Decisión: $F_{cal}=3,8465 > F_{5\%,6,24}=2,51$, por lo que se rechaza H_0, llegando a la conclusión de que existe evidencia de una diferencia entre los grupos de estimadores, de esta manera se ratifica la conclusión que la conformación de bloques ha sido ventajosa para reducir el error experimental.

7.4. ANAVA DE DOS FACTORES SIN REPLICACIÓN.

Los artículos fabricados por una compañía se producen por seis operarios utilizando ocho máquinas diferentes. El fabricante quiere determinar si hay diferencias significativas entre las máquinas y entre los operarios. Se efectúa un experimento para determinar el número de artículos diarios producidos por cada operario utilizando cada una de las máquinas, a un nivel de significación del 5% para probar si existe una diferencia significativa.

Datos.

Operarios	Máquinas							
	1	2	3	4	5	6	7	8
1	19,00	19,00	17,00	21,00	20,00	16,00	18,00	23,00
2	14,00	17,00	14,00	17,00	15,00	14,00	17,00	21,00
3	15,00	16,00	11,00	16,00	16,00	12,00	21,00	16,00
4	15,00	19,00	18,00	19,00	18,00	17,00	21,00	14,00
5	21,00	22,00	21,00	28,00	9,00	24,00	26,00	22,00
6	12,00	20,00	15,00	22,00	14,00	12,00	11,00	15,00

Solución.

Anava de dos factores sin replicación

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	Razón F calculada
Máquinas	156,00	7	22,2857	$F^C = 2,2018$
Operarios	253,75	5	50,7500	$F^F = 5,0141$
Error	354,25	35	10,1214	
Total	**764,00**	**47**		

Decisión: $F^C=2,2018 < F_{5\%,7,35}=2,29$ por lo que no se rechaza H_0^C, se concluye que no existe evidencia de una diferencia en entre las máquinas.

Decisión: $F^F=5,0141 > F_{5\%,5,35}=2,485$, por lo que se rechaza H_0^F, llegando a la conclusión de que existe evidencia de diferencia entre los operarios.

7.5. ANAVA DE DOS FACTORES CON REPLICACIÓN.

El hospital de una ciudad realiza un análisis del tiempo de visitas domiciliares individuales realizado por las enfermeras de salud pública, a cinco clases de enfermos, cuya duración es medida en minutos, además se registro la edad de cada una de las enfermeras; a partir de esta información se desea saber:

a) ¿El tiempo medio de visita difiere entre los distintos grupos de edad de las enfermeras?

b) ¿ El tipo de paciente influye en el tiempo medio de visita?

c) ¿Existe interacción entre las edades de las enfermeras y el tipo de pacientes?

Decisión: se rechaza H_0^C, se concluye que existe evidencia de una diferencia entre

Datos.

		Niveles del Factor B (Grupo de edades de las enfermeras)					
N.F.A.		**20-25**	**26-30**	**31-35**	**36-40**	**41-45**	**>=46**
		1	**2**	**3**	**4**	**5**	**6**
1		Cardiacos					
	1	22,00	23,00	26,00	30,00	23,00	27,00
	2	24,00	31,00	29,00	30,00	27,00	30,00
	3	23,00	30,00	23,00	25,00	25,00	27,00
	4	26,00	27,00	26,00	28,00	24,00	28,00
2		Cancerosos					
	1	29,00	31,00	35,00	38,00	37,00	41,00
	2	35,00	25,00	40,00	43,00	41,00	44,00
	3	32,00	33,00	38,00	45,00	35,00	49,00
	4	34,00	29,00	39,00	47,00	44,00	46,00
3		Sidosos					
	1	24,00	25,00	35,00	40,00	39,00	51,00
	2	31,00	25,00	35,00	36,00	43,00	44,00
	3	29,00	38,00	36,00	45,00	31,00	54,00
	4	35,00	25,00	32,00	31,00	35,00	30,00
4		Tuberculosos					
	1	32,00	34,00	29,00	31,00	29,00	27,00
	2	26,00	29,00	27,00	29,00	25,00	26,00
	3	21,00	18,00	19,00	27,00	24,00	31,00
	4	18,00	26,00	25,00	26,00	24,00	31,00
5		Neuróticos					
	1	31,00	32,00	29,00	28,00	34,00	37,00
	2	26,00	25,00	24,00	25,00	27,00	26,00
	3	19,00	28,00	26,00	30,00	24,00	31,00
	4	18,00	19,00	23,00	22,00	24,00	38,00

N. F. A. = Niveles del factor A (5 tipos de pacientes)

Solución.

Anava de dos factores con replicación.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrados Medios	Razón F calculada
Factor A Columnas	1.318,6720	5	263,7344	$F^{C} = 9{,}0969$
Factor B Filas	2.127,9530	3	709,7500	$F^{F} = 24{,}4662$
Interacción A*B	646,0938	15	43,0729	$F^{I} = 1{,}4867$
Error de Muestreo	2.783,2030	96	28,9917	
Total	**6.875,9220**	**119**		

las edades de las enfermeras, por lo tanto, el tiempo medio de visita difiere entre los distintos grupos de edad de las enfermeras.

Decisión: F^{F}=9,0969 > $F_{5\%,3,96}$=8,5616 por lo que no se rechaza H_0^{F}, se concluye que existe evidencia de una diferencia entre los tipos de pacientes.

Decisión: F^{I}=1,4857 < $F_{5\%,15,96}$=1,775 , por lo que se rechaza H_0^{I}, llegando a la conclusión de que existe evidencia de una interacción entre las edades de las enfermeras y el tipo de pacientes.

7.6. ANAVA CON N FACTORES.

Esta aplicación se la efectuará por medio de un ejemplo, en el que se considera un experimento factorial con cuatro niveles A,B,C y R en un diseño por bloques completamente aleatorizados, como se presenta en la tabla siguiente:

X_{abcr} donde: a=1,2,3,4,5 b=1,2,3 c=1,2,3,4
r = 1,2

Solución. Ejecutando esta opción, se tienen los datos de salida de la siguiente manera:

Tabla 5.
Datos de la muestra para el ANAVA con cuatro factores

		b_1					b_2					b_3				
Replica del bloque		a_1	a_2	a_3	a_4	a_5	a_1	a_2	a_3	a_4	a_5	a_1	a_2	a_3	a_4	a_5
r_1	c_1	7	20	12	7	14	8	19	9	22	18	10	9	13	11	10
	c_2	8	22	13	8	12	7	17	12	23	12	11	8	13	14	9
	c_3	9	20	15	9	13	9	16	13	24	18	9	7	15	17	15
	c_4	8	24	19	10	19	10	15	14	22	17	8	8	12	16	9
r_2	c_1	9	21	15	11	18	11	17	8	26	16	10	9	17	14	17
	c_2	8	20	17	10	18	12	16	11	19	19	12	10	19	15	18
	c_3	11	20	18	18	18	13	14	15	24	18	18	14	19	15	19
	c_4	12	22	21	15	17	15	17	16	25	19	17	15	21	19	20

En este experimento el factor A tienen cinco niveles, el factores B tiene tres niveles y C tiene cuatro niveles y el experimento es replicado dos veces. Las replicaciones son completamente relacionadas y no constituyen un factor. Sin embargo, para los propósitos que se persigue se lo considera a R como un cuarto factor asumido. De esta manera, cada elemento de los datos en la Tabla 5 es representado en esta forma:

Niveles de los factores	
A	5
B	3
C	4
R	2

Gran media= 14,68333

Niveles			Nivel 1				
4	3	2					
1	1	1	7	20	12	7	14
1	1	2	8	19	9	22	18
1	1	3	10	9	13	11	10
.	.	.	..	..	..	..	..
2	4	1	12	22	21	15	17
2	4	2	15	15	16	25	19
2	4	3	17	15	21	19	20

Solución.

Anava con n factores (n=4)

Anava con 4 factores			
Fuente de Variación	**Suma de Cuadrados**	**Grados de Libertad**	**Cuadrados medios**
A	552,55000	4	138,13750
B	110,46670	2	55,23334
AB	1.141,20000	8	142,65000
C	126,43330	3	42,14445
AC	50,98334	12	4,24861
BC	17,46667	6	2,91111
ABC	77,86667	24	3,24444
R	240,83330	1	240,83330
AR	45,45667	4	11,35417
BR	66,46667	2	33,23333
ABR	62,03334	8	7,75417
CR	15,76667	3	5,25556
ACR	51,31667	12	4,27639
BCR	32,93334	6	5,48889
ABCR	70,23333	24	2,92639
Total	**2.661,96700**	**119**	

En nuestro país, otra importante y útil aplicación del diseño factorial es en Ingeniería de Minas, en cuanto se refiere a los tratamientos efectuados en la flotación de minerales, como en el caso particular de la recuperación del Sn contenido en el mineral casiterita (SnO_2) que se presenta con otros minerales. Los factores que se utilizan y sus niveles más próximos pueden ser: tiempo de flotación (5, 10 o 15 minutos,), grado de acidez (2, 3 o 4 pH), ácido cítrico (0, 5, o 10 cm^3) y aereosol 22 (10, 20 o 30 cm^3) por celda de flotación, valores más convenientes de los niveles los determina el ingeniero, con el objeto de tener la mejor recuperación de Sn, por consiguiente el mejor rendimiento económico.

Colaboración.

Prof. M. Nilda Avilés de Ruiz

Lic. en Idiomas. Universidad Autónoma Gabriel René Moreno

Santa Cruz – Bolivia (Agosto, 2005)

M. Cs. Astrid Keitel Avilés

Lic. en Adm. Emp. Universidad Católica Boliviana, Santa Cruz – Bolivia (Octubre,1999)

M. Cs. Academia Diplomática Boliviana ¨Rafael Bustillo¨, La Paz – Bolivia (Abril, 2004)

BIBLIOGRAFÍA

BERENSON L. Mark, LEVINE M. David, "*Estadística Básica en Administración*". Prentice Hall Hispanoamericana, S.A., México D.F. 1997 (6ta. edición), pp. 943 - XIX

COOLEY W. William, LOHNES R. Paul, "*Multivariate Data Analysis*". John Wiley & Sonns Inc., New York, Estados Unidos, 1971 (1ra edición), pp. 364 – XIII

CORDOVA Zamora Manuel, "*Estadística Inferencial*", Editorial Moshera, Lima, Perú, 1999 (1ra edición), pp. 410 – XV

DAGNELIE Pierre, "*Analyse Statistique á Plusieurs Variables*". Les Presses Agronomiques de Gembloux, Gembloux, Belgica, 1975 (2da. Edición), pp. 362 - XIV

International Business Machines Corporation, System Reference Library, "*Scientific Subroutines*", New York, USA, 1967, pp. 191.

JOHNSON A. Richard, "*Probabilidad y Estadística para Ingenieros*". Prentice Hall Hispanoamericana, S.A., México D.F. 1997 (5ta. Edición), pp. 630 - XV

LEBART Ludovic, FÉNELON Pierre, "*Statistique et Informatique Appliquées*", Dunod, París, Francia, 1977 (3ra edición), pp. 439 – V.

RUIZ Aranibar Gustavo. *Factores que Inciden em El Rendimiento Académico y Evaluación Docente*". U.A.G.R.M. Santa Cruz – Bolivia. 2010, pp. 190 – IX.

RUIZ Aranibar Gustavo[2] "*Libreria Científica de Programas Informáticos*", La Paz -Bolivia.

WAYNE Daniel, W. "*Bioestadística*", Uthea, México D.F. 1999 (1ra. Edición), pp. 878 – XII

WINER B. J. "*Statistical Principle in Experimental Design*", McGraw Hill, New York, USA, 1971, (2da. Edición), pp. 907 - X.

Para triunfar, toda la vida debemos prepararnos, estudiar, aprender, transmitir los conocimientos adquiridos, lo cual permite que el estudiante para ser un buen profesional, debe tratar de ser mejor que sus docentes.

Gustavo Ruiz Aranibar

[2] Calle 20 y Av. Ballivian, Nº 8035, Calacoto, La Paz – Bolivia, Tel. 591-22772162 Cel, 67111778
gustavoruiz432@hotmail.com.bo *ruizaranibargustavo@gmail.com.bo* Blog: Gustavo Ruiz Ar