Diseños factoriales

Diplomado: Diseño Profesional de Experimentos
Modulo: Fundamentos del diseño experimental

Econ. Morales Alberto Alexis Adonai

2025

Índice

Índice

Índice

Índice

Índice

Índice

Objetivos

- Describir los conceptos básicos en diseños factoriales y revisar los detalles de cómo se aplican.
- Estudiar los diseños factoriales de dos y tres factores, además de la mandra en la que estabilizan la varianza.
- Explicar el diseño factorial de manera general, el modelo de efectos fijos y la diferencia con el modelo de efectos aleatorios.

Sección 1

Conceptos básicos en diseños factoriales

Conceptos básicos en diseños factoriales

- El **diseño factorial** es un diseño experimental que sirve para estudiar el efecto individual y de interacción de varios factores (a partir de 2) sobre una o varias variable(s) de respuesta.
- Los **factores** pueden ser de tipo *cualitativo* (máquinas, tipos de material, operador, presencia o ausencia de una operación previa), o de tipo *cuantitativo* (temperatura, humedad, velocidad, presión, etc).
- Para estudiar la forma en que influye cada factor sobre la variable de respuesta, será necesario elegir al menos dos niveles de prueba para cada uno de ellos. Con el diseño factorial completo, se corren aleatoriamente todas las posibles combinaciones que pueden formarse en los niveles de los factores.

Conceptos básicos en diseños factoriales

- La **matriz** de diseño o *arreglo* factorial, es el conjunto de puntos experimentales o tratamientos que pueden formarse al considerar todas las posibilidades d combinación de los niveles de los factores.
- El nombre del diseño factorial va implícito en el número de tratamientos que lo componen.
- Para obtener el número de corridas experimentales, se multiplica el número de tratamientos por el número de *replicas*, donde una de éstas se lleva a cabo en vez que se corre el arreglo completo.
- En general, la familia de diseños factoriales 2^k consiste en k factores, todos con dos niveles de prueba; y la familia de diseños factoriales 3^k, consiste en k factores, cada uno con tres niveles de prueba.

Conceptos básicos en diseños factoriales

- Es claro que si los k factores no tienen la misma cantidad de niveles, debe escribirse el producto de manera explícita, por ejemplo, con $k = 3$ factores, el primero con cuatro niveles y los dos restantes con dos niveles, se tiene un diseño factorial $4 \times 2 \times 2$ ó 4×2^2.

Conceptos básicos en diseños factoriales

- **Efecto de un factor**.- Es el cambio observado en una variable de respuesta debido a un cambio de niveles en el factor.
- **Efecto principal**.- Es igual a la respuesta promedio observada en el nivel alto de un factor, menos la respuesta promedio en el nivel bajo.
- **Efecto de interacción**.- Dos factores interactúan de manera significativa sobre la variable de respuesta cuando el efecto de uno depende del nivel en que está el otro.

Para poder encontrar que los efectos son estadísticamente significativos (diferentes de cero), se requiere el análisis de varianza (ANOVA).

Representación de los efectos principales y la interacción

- El efecto principal de un factor se representa de manera gráfica, en cuyo eje horizontal se ubican los niveles del factor y en el eje vertical se encuentra la media de la respuesta observada en los correspondientes niveles.

Ejemplo de intereacción de efectos

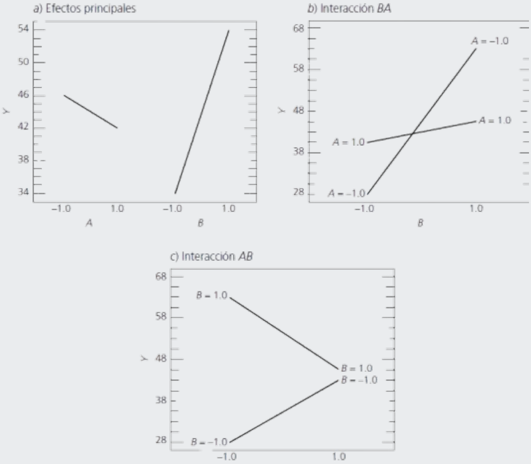

Figura 1: Interacción de los efectos principales

Ejemplo de intereacción de efectos

- En el ejemplo anterior, cuando hay interacción de los factores, las línea tienen una pendiente muy diferente (inciso b), ya que se muestra la interacción AB, poniendo el factor B en el eje horizontal.
- En el inciso C, se representa el mismo efecto que em B, solo que ahora con el factor A en el eje horizontal.
- Para entender concretamente el ejemplo, se debe ser cuidadoso y analizar con detalle lo que pasa en Y cuando se ve un factor dependiendo del nivel, por ejemplo, en el inciso C, se aprecia que si A cambia de nivel (-1) a (1), cuando B = -1, la respuesta Y también se incrementa; pero si B = 1, la respuesta decrece de manera importante.

Ejemplo de intereacción de efectos

- El factor A, tiene un efecto positivo o negativo en Y, dependiendo del nivel de B.
- En el caso del inciso B, se puede ver que si B incrementa (cambio de -1 a 1) cuando A = 1, la respuesta Y se incrementa ligeramente, pero si A = -1, la respuesta incrementa mucho.

Ejemplo de mo intereacción de efectos

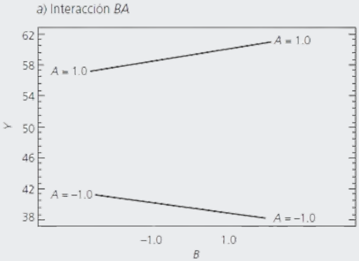

Figura 2: No interacción de efectos principales

Ejemplo de no interacción de efectos

- En la figura anterior, en el inciso a) se ve que el factor B aumenta, pero la respuesta de Y prácticamente no cambia en ambos niveles de A. Por lo tanto, no hay interacción entre los factores.
- Lo mismo sucede en el inciso b), se ve que el facto A aumenta y la variable de respuesta aumenta, pero sin importar el valor de B, ya que no se nota diferencia en el cambio de A cuando B cambia y el nivel de Y se mantiene similar.

Sección 2

Experimentación factorial vs mover un factor a la vez

Experimentación factorial vs mover un factor a la vez

- Los diseños factoriales son más eficientes que el tradicional experimento de mover un factor a la vez, que utilizan las personas cuando no tienen conocimiento del diseño de experimentos.
- Una forma de ver la ineficiencia de mover un factor a la vez, se ilustra considerando el siguiente ejemplo.
- Se trata de estudiar los efectos sobre el rendimiento de un proceso que tienen tres factores: A (tempratura); B (contenido de sólidos) y C (tiempo de residencia). Cada factor se estudiará a dos niveles (-, +).

Experimentación factorial vs mover un factor a la vez

- De acuerdo con el enfoque de experimentación de mover un factor a la vez, se procede de la siguiente forma:
  1. Para estudiar el efecto de A se realizan cuatro pruebas con cada nivel de A, mientras que los factores B y C se fijan. Se obtiene que A_+ es mejor que A_-.
  2. Ahora se hace lo mismo para el factor B, pero fijando A en (+) que fue el mejor y a C en (-). Con las cuatro pruebas de cada nivel, se obtiene que con B_+ se logra un mejor rendimiento que con B_-.
  3. De acuerdo con lo anterior se fija A_+ y B_+ y de igual manera se estudia el efecto de C. Se obtiene que C_+ es mejor que C_-,
  4. Conclusión: Condición optima (A_+, B_+, C_+)

Experimentación factorial vs mover un factor a la vez

Tratamiento	A	B	C	Paso 1	Paso 2	Paso 3	Total de corridas
1	-	-	-	*			4
2	+	-	-	*	*		8
3	-	+	-				0
4	+	+	-		*	*	8
5	-	-	+				0
6	+	-	+				0
7	-	+	+				0
8	+	+	+			*	4
Total							**24**

Cuadro 1: Experimentación empírica

Experimentación factorial vs mover un factor a la vez (problema)

- La mejor condición que maximiza el rendimiento puede ser cualquiera de los cuatro tratamientos que no se probaron, como se mostró en el *cuadro 1*, en la que se aprecian las ocho combinaciones o tratamientos diferentes que resultaron de tener tres factores, cada uno con dos niveles.
- Los tratamientos que sí se probaron en cada paso y el total de corridas en las que se realizó cada tratamiento, hay tratamientos que sin saberlo fueron probados en dos pasos (ocho veces), y otros en ninguno.
- En consecuencia, después de realizar 24 pruebas sólo se estudiaron la mitad de los posibles tratamientos.

Experimentación factorial vs mover un factor a la vez (problema)

- La comparación entre los resultados de un tratamiento y otro no se hizo con un criterio estadístico, y no se estudio el efecto simultáneo de los factores (interacción), por lo tanto, se gastaron muchos recursos y se ha obtenido poca información y no garantiza que la solución propuesta sea la mejor.

Experimentación factorial vs mover un factor a la vez (enfoque correcto)

- Es mejor aplicar un diseño factorial, con el cual se investigan (en orden aleatorio) las ocho combinaciones de los niveles de los factores, en el caso descrito antes, con 24 pruebas o incluso menos, 16 por ejemplo, se podrían haber estudiado los ocho tratamientos con dos o tres replicas, y después de un análisis estadístico adecuado, saber qué factores son más importantes, si interactuan sobre Y u concluir cuál es el tratamiento que más conviene.
- La experimentación empírica en apariencia más simple y sencilla ha resultado cara y poco eficaz.

Ventajas de los diseños factoriales

1. Permiten estudiar el efecto individual y la interacción de los factores.
2. Son diseños que se pueden aumentar para formar diseños compuestos en caso de que se requiera una exploración más completa.
3. Se pueden correr fracciones de diseños factoriales, las cuales son de gran utilidad en las primeras etapas de un análisis que involucra a muchos factores, cuando interesa descartar los que no son importantes.

Ventajas de los diseños factoriales

4. Pueden utilizarse en combinación con diseño de bloques en situaciones en las que no puede correrse todo el diseño factorial bajo las mismas condiciones.
5. La interpretación y el cálculo de los efectos en los experimentos factoriales se puede hacer con aritmética elemental, en particular cuando cada factor se prueba en dos niveles.

Sección 3

Diseños factoriales con dos factores

Diseños factoriales con dos factores

- Consideremos los factores A y B con a y b $(a, b \geq 2)$ niveles de prueba, respectivamente. Con ellos se puede construir el arreglo o diseño factorial $a \times b$, el cual consiste en $a \times b$ tratamientos. Algunos casos particulares de uso frecuente son: el factorial 2^2, el factorial 3^2 y el factorial 3×2.
- **Replica**.- Es cada corrida de todos los tratamientos del arreglo factorial.
- Los diseños factoriales que involucran menos de cuatro factores por lo regular se corrren replicados para tener la potencia necesaria en las pruebas estadísticas sobre los efectos de interés. En caso de realizar n replicas, el número total de corridas experimentales es $n\,(a \times b)$.

Modelo estadístico e hipótesis de interés

- El modelo estadístico de efectos para este tipo de diseño está dado por:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{1}$$

$$i = 1, 2, \ldots, a; \quad j = 1, 2, \ldots, b; \quad k = 1, 2, \ldots, n$$

Modelo estadístico e hipótesis de interés

- Donde μ es la media general, α_i es el efecto debido al i-ésimo nivel del factor A, β_j es el efecto del j-ésimo nivel del factor B, $(\alpha\beta)_{ij}$ representa el efecto de interacción en la combinación ij, y ϵ_{ijk} es el error aleatorio que, se supone, sigue una distribución normal con media cero y varianza constante $\left(N(0, \sigma^2)\right)$ y son independientes entre sí.
- Para que la estimación de los parámetros en este modelo sea única, se introducen las restricciones: $\sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0 \quad$ y $\quad \sum_{i=1}^{a} \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0.$. Los efectos dados en el modelo son desviaciones relacionadas con la media global.

Planteamiento de hipótesis

- Las hipótesis planteadas con los efectos descritos para el modelo (1) son las siguientes:

$$
\begin{gathered}
H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \\
H_A : \alpha_i \neq 0 \text{ para algún } i \\[1ex]
H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0 \\
H_A : \beta_j \neq 0 \text{ para algún } j \\[1ex]
H_0 : (\alpha\beta)_{ij} = 0 \text{ para todo } i, j \\
H_A : (\alpha\beta)_{ij} \neq 0 \text{ para algún } i, j
\end{gathered} \tag{1.H}
$$

Diseños factoriales con dos factores

- Estad hipótesis se prueban mediante la técnica de análisis de varianza, que para un diseño factorial $(a \times b)$ con n réplicas resulta de descomponer la variación total como:

$$SC_T = SC_A + SC_B + SC_{AB} + SC_E \tag{2}$$

- Donde los respectivos grados de libertad de cada una de ella son:

$$nab - 1 = (a-1) + (b-1) + (a-1)(b-1) + ab(n-1) \tag{2.1}$$

Diseños factoriales con dos factores

- El factor $(n-1)$ en los grados de libertad de la suma de cuadrados del error (SC_E) señala que se necesitan al menos dos réplicas del experimento para calcular este componente y, por ende, para construir una tabla ANOVA.
- La suma de cuadrados divididas entre sus correspondientes grados de libertad se llama *cuadrados medios* (CM). Al dividir éstos entre el cuadrado medio del error CM_E se obtienen estadísticos de prueba con distribución F.

Diseños factoriales con dos factores

Fuente de variabilidad	Suma de cuadrados	gl	Cuadrado medio	F_0	Valor-p
Efecto A	SC_A	$a-1$	CM_A	$\frac{CM_A}{CM_E}$	$P(F > F_0^A)$
Efecto B	SC_B	$b-1$	CM_B	$\frac{CM_B}{CM_E}$	$P(F > F_0^B)$
Interacción AB	SC_{AB}	$(a-1)(b-1)$	CM_{AB}	$\frac{CM_{AB}}{CM_E}$	$P(F > F_0^{AB})$
Error	SC_E	$ab(n-1)$	CM_E	–	–
Total	SC_T	$abn-1$	–	–	–

Cuadro 2: ANOVA para el diseño factorial a x b

Diseños factoriales con dos factores

- Si el valor-p es menor que el nivel de significancia α prefijado, se rechaza la hipótesis nula y se concluye que el correspondiente efecto está activo o influye en la variable de respuesta.
- Las siguientes fórmulas derivan al calculo de SC_T

$$Y_{***} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk} \quad \overline{Y}_{***} = \frac{Y_{***}}{abn}$$

$$Y_{i**} = \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk} \quad \overline{Y}_{i**} = \frac{Y_{i**}}{bn} \quad i = 1, 2, \dots, a$$

$$Y_{*j*} = \sum_{i=1}^{a} \sum_{k=1}^{n} Y_{ijk} \quad \overline{Y}_{*j*} = \frac{Y_{*j*}}{an} \quad j = 1, 2, \dots, b$$

$$Y_{ij*} = \sum_{k=1}^{n} Y_{ijk} \quad \overline{Y}_{ij*} = \frac{Y_{ij*}}{n}$$

- Con la notación anterior, la suma de cuadrados totales es:

$$SC_T = \sum_{i=1}^{a} Y_{ijk}^2 - \frac{Y_{***}^2}{N}$$

- Donde $N = abn$ es el total de observaciones en el experimento.

- Las sumas de los cuadrados son:

$$SC_A = \sum_{i=1}^{a} \frac{Y_{i**}^2}{bn} - \frac{Y_{***}^2}{N}$$

$$SC_B = \sum_{j=1}^{b} \frac{Y_{*j*}^2}{an} - \frac{Y_{***}^2}{N}$$

$$SC_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{Y_{ij*}^2}{n} - \frac{Y_{***}^2}{N} - SC_A - SC_B$$

Diseños factoriales con dos factores

- Al final, al restar todas las sumatorias, se obtiene la suma de cuadrados del error como:

$$SC_E = SC_T - SC_A - SC_B - SC_{AB}$$

Ejemplo de diseño factorial con dos factores

- Considerando lo anterior, se reeplicara el proceso con el siguiente ejemplo:

Cuadro 3: Datos del experimento factorial 4×3

A: profundidad	B: velocidad			Total $Y_{i\cdot\cdot}$
	0.20	0.25	0.30	
0.15	74	92	99	
	64	86	98	763
	60	88	102	
0.18	79	98	104	
	68	104	99	808
	73	88	95	
0.21	82	99	108	
	88	108	110	881
	92	95	99	
0.24	99	104	114	
	104	110	111	944
	96	99	107	
Total $Y_{\cdot j\cdot}$	**979**	**1 171**	**1 246**	$Y_{\cdots} = 3\,396$

Ejemplo de diseño factorial con dos factores

- Siguiento el cuadro 3, el proceso se calcula de la siguiente manera:

$$SC_A = \sum_{i=1}^{4} \frac{Y_{i**}^2}{3 \times 3} - \frac{Y_{***}^2}{4 \times 3 \times 3} = \frac{763^2 + 808^2 + 881^2 + 944^2}{3 \times 3} - \frac{3396^2}{4 \times 3 \times 3} = 2125.1$$

$$SC_B = \sum_{j=1}^{3} \frac{y_{*j*}^2}{4 \times 3} - \frac{Y_{***}^2}{4 \times 3 \times 3} = \frac{979^2 + 1171^2 + 1246^2}{4 \times 3} - \frac{3396^2}{4 \times 3 \times 3} = 3160.5$$

$$SC_{AB} = \sum_{j=1}^{4} \sum_{i=1}^{3} \frac{y_{ij*}}{3} - \frac{y_{***}^2}{4 \times 3 \times 3} - SC_A - SC_B =$$

$$\frac{198^2 + 220^2 + \cdots + 332^2}{3} - \frac{3396^2}{4 \times 3 \times 3} - 2125.1 - 3160.5 = 557.07$$

Ejemplo de diseño factorial con dos factores

- La suma de cuadrados totales y la suma de cuadrados del error están dadas por:

$$SC_T = \sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{l=1}^{3} Y_{ijl}^2 - \frac{Y_{***}^2}{4 \times 3 \times 3} = 6532.0$$

$$SC_E = SC_T - SC_A - SC_B - SC_{AB} = 6532.0 - 2125.1 - 3160.5 - 557.07 = 689.33$$

- Con ello se construye la salida de la tabla ANOVA

Ejemplo de diseño factorial con dos factores

Cuadro 4: Análisis de varianza (ANOVA) para el experimento factorial

Fuente de variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor-p
A: prof	2125.10	3	708.37	24.66	0.0000
B: veloc	3160.50	2	1580.25	55.02	0.0000
AB	557.07	6	92.84	3.23	0.0180
Error	689.33	24	28.72		
Total	6532.00	35			

Diseños factoriales con dos factores

- Con la información del cuadro 4 se construye el análisis de varianza, de donde se concluye que los tres efectos A (Velocidad), B (profundidad) y AB están activos o influyen en el acabado.
- Dado que el efecto de interacción AB resulta significativo, prácticamente toda la información relevante del experimento se aprecia en su presentación gráfica.

Diseños factoriales con dos factores

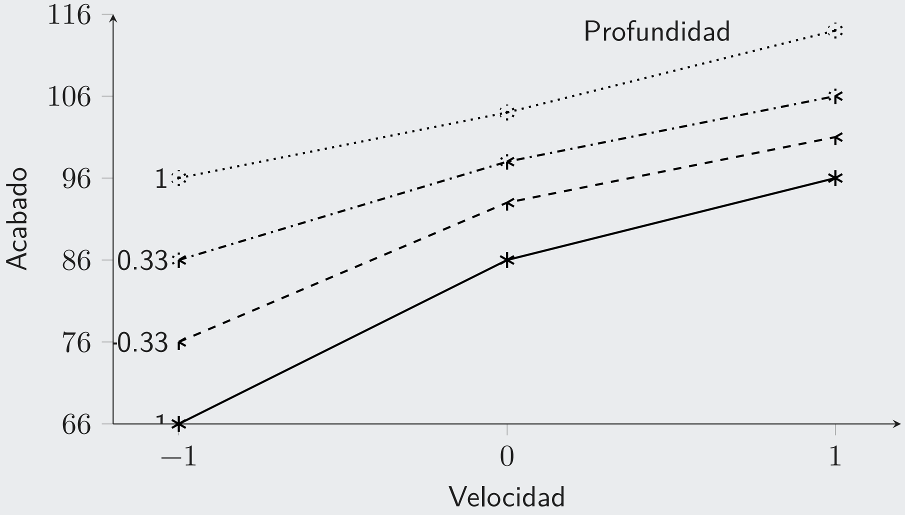

Figura 3: Efecto de interacción velocidad × profundidad

Diseños factoriales con dos factores

- Hay que notar en la figura 3, que aparecen tantas lineas como niveles tenga el factor, que se dibuja en la parte de arriba, que en este caso es la profundidad con sus cuatro niveles que denotan con la escala de -1 a 1.
- La significancia de la interacción detectada por el ANOVA se observa en el hecho de que las líneas tienen pendientes relativamente diferentes.
- Como interesa minimizar la variable de respuesta, se observa que a mayor velocidad y profundidad hay una tendencia a obtener peores acabados.
- Se ve que cuando tiene velocidad alta (A^+) el efecto de profundidad es menor. Por lo tanto las condiciones de operación o tratamiento que convienen en profundidad y velocidad bajas (A^-, B^-)

Comparación de medias

Las *comparaciones de medias* se introducen para investigar, después de un ANOVA (pruebas post hoc) en el que rechaza H_0, cuáles medidas causan las diferencias detectadas. El ANOVA sólo indica que al menos un par de niveles del factor significativo son diferentes entre sí, pero no dice cuáles son.

- Para facilidad, denotemos los cuatro niveles de profundidad (A) del ejemplo que se vio anteriormente, como A_1, A_2, A_3 y A_4, así como los tres niveles de la velocidad (B) como B_1, B_2 y B_3.
- Los seis pares de hipótesis para comparar las medias del factor A son las siguientes.

Comparación de medias

$$
\begin{array}{lll}
H_0: \mu_{A_1} = \mu_{A_2} & H_0: \mu_{A_1} = \mu_{A_3} & H_0: \mu_{A_1} = \mu_{A_4} \\
H_A: \mu_{A_1} \neq \mu_{A_2} & H_A: \mu_{A_1} \neq \mu_{A_3} & H_A: \mu_{A_1} \neq \mu_{A_4} \\
H_0: \mu_{A_2} = \mu_{A_3} & H_0: \mu_{A_2} = \mu_{A_4} & H_0: \mu_{A_3} = \mu_{A_4} \\
H_A: \mu_{A_2} \neq \mu_{A_3} & H_A: \mu_{A_2} \neq \mu_{A_4} & H_A: \mu_{A_3} \neq \mu_{A_4}
\end{array}
$$

Comparación de medias

- Mientras que para el factor B, se tienen los tres pares de hipótesis:

$$\begin{array}{lll} H_0\colon \mu_{B_1} = \mu_{B_2} & H_0\colon \mu_{B_1} = \mu_{B_3} & H_0\colon \mu_{B_2} = \mu_{B_3} \\ H_A\colon \mu_{B_1} \neq \mu_{B_2} & H_A\colon \mu_{B_1} \neq \mu_{A_3} & H_A\colon \mu_{B_2} \neq \mu_{B_3} \end{array}$$

Comparación de medias

- Para probar estas hipótesis se utiliza el método LSD, con el cuál habrá que calcular las diferencias muestrales en valor absoluto y comprarlas con la diferencia mínima signficativa.
- Cabe resaltar que este análisis es engañoso cuando el efecto de interacción es significativo, para el ejemplo elaborado, se ignorará ello y solo se probarán las hipótesis del factor A.

- La diferencia mínima significativa para comparar los niveles i y l del factor A, esta dada por:

$$LSD_A = t_{\alpha/2,ab(n-1)}\sqrt{CM_E\left(\frac{1}{n_{A_i}}+\frac{1}{n_{A_l}}\right)}$$

- Donde $t_{\alpha/2,ab(n-1)}$ es el punto porcentual $100\,(1-\alpha/2)$ de la distribución T de Student, $ab\,(n-1)$ son los grados de libertad del cuadrado medio del error, n_{A_i} y n_{A_l} son el total de observaciones en los niveles i y l del factor A.

Comparación de medias

- Como el ejemplo que se ha trabajado, es un diseño balanceado $n_{A_i} = n_{A_l} = 9$, entonces:

$$LSD_A = 2.064\sqrt{28.72\,(2/9)} = 5.21$$

- De los totales marginales, se obtienen las medias del factor A, al dividir entre 9, que son el número de mediciones involucradas en cada total. Así, las seis posibles diferencias muestrales en valor absoluto resultan ser las siguientes:

Comparación de medias

$$\left|\bar{Y}_{A_1} - \bar{Y}_{A_2}\right| = \frac{1}{9}\left|763 - 808\right| = 5.0 < LSD_A$$

$$\left|\bar{Y}_{A_1} - \bar{Y}_{A_3}\right| = \frac{1}{9}\left|763 - 881\right| = 13.1^{*} > LSD_A$$

$$\left|\bar{Y}_{A_1} - \bar{Y}_{A_4}\right| = \frac{1}{9}\left|763 - 944\right| = 20.1^{*} > LSD_A$$

$$\left|\bar{Y}_{A_2} - \bar{Y}_{A_3}\right| = \frac{1}{9}\left|808 - 881\right| = 8.1^{*} > LSD_A$$

$$\left|\bar{Y}_{A_2} - \bar{Y}_{A_4}\right| = \frac{1}{9}\left|808 - 944\right| = 15.1^{*} > LSD_A$$

$$\left|\bar{Y}_{A_3} - \bar{Y}_{A_4}\right| = \frac{1}{9}\left|881 - 944\right| = 7.0^{*} >< LSD_A$$

Comparación de medias

- Sólo la primer diferencia resulta no significativa, es decir, se acepta $H_0 : \mu_{A_1} = \mu_{A_2}$, en cambio, en las cinco comparaciones restantes, se rechaza H_0.

Comparación de medias: Tomando en cuenta la interacción

- Para hacer comparaciones múltiples de medidas de un factor, tomando en cuenta el efecto de interacción, éstas se deben realizar de manera separada en cada nivel del otro factor. En el ejemplo que se ha estado mostrando, las comparaciones que se hicieron para el factor A se realizan dentro de cada nivel del factor B, de esta forma, se tomará en cuenta el efecto de interacción y por ende se interpretará más cercano a la realidad.

Comparación de medias: Tomando en cuenta la interacción

- Si se observa el efecto de la figura 3, es fácil notar que las medias de las tres últimas profundidades (denotadas -0.33, 0.33 y 1) están cercanas entre sí cuando la velocidad está en su nivel intermedio que cuando está en su nivel bajo.
- Hay que verlo de manera analitica en la velocidad intermedia (B_2), donde las medias muestrales del factor A (profundidad) en la velocidad intermedia son:

$$\bar{Y}_{1,2,\cdot} = \frac{266}{3} = 88.66$$
$$\bar{Y}_{2,2,\cdot} = \frac{290}{3} = 96.66$$
$$\bar{Y}_{3,2,\cdot} = \frac{302}{3} = 100.66$$
$$\bar{Y}_{4,2,\cdot} = \frac{313}{3} = 104.33$$

Comparación de medias: Tomando en cuenta la interacción

- Entonces, para comparar estas medias la diferenci mínima significativa está dada por:

$$LSD_{B_2(A)} = t_{\alpha/2,ab(n-1)}\sqrt{CM_E\left(\frac{1}{n}+\frac{1}{n}\right)}$$

- Donde n es el número de réplicas de los tratamientos a comparar.

$$LSD_{B_2(A)} = 2.064\sqrt{28.72\left(\frac{2}{3}\right)} = 9.03$$

Comparación de medias: Tomando en cuenta la interacción

- Al comparar las seis posibles diferencias de medias de los niveles de A en valor absoluto contra la cota $LSD_{B_2(A)}$.

$$\left|\bar{Y}_{B_2(A_1)} - \bar{Y}_{B_2(A_2)}\right| = |88.66 - 96.66| = 8.0 < LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_1)} - \bar{Y}_{B_2(A_3)}\right| = |88.66 - 100.66| = 12.0^{*} > LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_1)} - \bar{Y}_{B_2(A_4)}\right| = |88.66 - 104.33| = 15.7^{*} > LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_2)} - \bar{Y}_{B_2(A_3)}\right| = |96.66 - 100.66| = 4.0 < LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_2)} - \bar{Y}_{B_2(A_4)}\right| = |96.66 - 104.33| = 7.7 < LSD_{B_2(A)}$$
$$\left|\bar{Y}_{B_2(A_3)} - \bar{Y}_{B_2(A_4)}\right| = |100.66 - 104.33| = 3.7 < LSD_{B_2(A)}$$

Comparación de medias: Tomando en cuenta la interacción

- Al tomar en cuenta el efecto de interacción AB se concluye que, cuando $B = B_2$, sólo hay diferencias entre el nivel A_1, con A_3 y A_4.
- Cuando los factores son cuantitativos, es importante verificar la posible presencia de curvatura en las diferentes gráficas de efectos: de interacción, de medias y de dispersión.
- La curvatura se evalúa analíticamente con el ANOVA desglosado.
- Las representaciones gráficas de efectos principales y de interacción significativos son suficientes para elegir el mejor tratamiento del esperimento.

Verificación de supuestos

- No hay que olvidar, que el supuesto de *normalidad*, *varianza constante* e *independencia de los residuos* en un diseño factorial se verifican principalmente con los métodos presentados en los temas pasados.

Sección 4

Diseños factoriales con tres factores

Diseños factoriales con tres factores

- Cuando se quiere investigar la influencia de tres factores (A, B y C) sobre una o más variables de respuesta, y el número de niveles de prueba de cada uno de los factores es a, b y c, respectivamente, se puede construir el arreglo factorial $a \times b \times c$, que consiste de $a \times b \times c$ tratamientos o puntos experimentales
- Los arreglos de este tipo que se utilizan con frecuencia en aplicaciones diverasas se encuentran: el factorial 2^3, el factorial 3^3 y los factoriales mixtos con no más de cuatro niveles en dos de los factores.

Modelo estadístico

- El diseño factorial $a \times b \times c$, se supone que el comportamiento de la respuesta Y puede describirse mediante el modelo de efectos dado por:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl} \quad (3)$$
$$i = 1, 2, \cdots, a;\ j = 1, 2, \cdots, b;\ k = 1, 2, \cdots, c;\ l = 1, 2, \cdots, n$$

Modelo estadístico

En la expresión (3); μ es la media general, α_i es el efecto del nivel i-ésimo del factor A, β_j es el efecto del nivel j del factor β y γ_k es el efecto del nivel k en el factor C; $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ y $(\beta\gamma)_{jk}$ representan efectos de interacción dobles (de dos factores) en los niveles ij, ik, jk, respectivamente, y $(\alpha\beta\gamma)_{ijk}$ es el efecto de interacción triple en la combinación o punto ijk; ϵ_{ijkl} representa el error aleatorio en la combinación $ijkl$ y l son las repeticiones o réplicas del experimento.

- Todos los efectos cumplen la restricción de sumar cero, es decir, son desviaciones relacionadas con la media general μ.

Hipótesis de interés

- El estudio factorial de tres factores (A, B y C) permite investigar los efectos A, B, C, AB, AC, BC y ABC, donde el nivel de desglose o detalle con el que pueden estudiarse depende del número de niveles utilizado en cada factor.
- De esta manera, se tiene siete efectos de interés sin considerar desglose, y con ellos se pueden plantear siete hipótesios nulas H_0 : Efecto $A = 0$, H_0 : Efecto $B = 0$, $\cdots$, H_0 : Efecto $ABC = 0$, cada una aparejada con su correspondiente hipótesis alternativa.
- El ANOVA para probar estas hipótesis se muestra en la siguiente tabla.

Estructura de ANOVA

Cuadro 5: Tabla de análisis de varianza trifactorial (ANOVA)

Fuente de variabilidad	**Suma de cuadrados**	**Grados de libertad**	**Cuadrado medio**	**F**	**Valor-p**
Efecto A	SCA	$a-1$	CMA	CMA/CME	$P(F > F_A)$
Efecto B	SCB	$b-1$	CMB	CMB/CME	$P(F > F_B)$
Efecto C	SCc	$c-1$	CMc	CMc/CME	$P(F > F_C)$
Efecto AB	SCAB	$(a-1)(b-1)$	CMAB	CMAB/CME	$P(F > F_{AB})$
Efecto AC	SCAC	$(a-1)(c-1)$	CMAC	CMAC/CME	$P(F > F_{AC})$
Efecto BC	SCBC	$(b-1)(c-1)$	CMC	CMC/CME	$P(F > F_{BC})$
Efecto ABC	SCABC	$(a-1)(b-1)(c-1)$	CMABC	CMABC/CME	$P(F > F_{ABC})$
Error	SCE	$abc(n-1)$	CME		
Total	SCT	$abcn-1$			

Estructura de ANOVA

- El efecto cuyo valor-p sea menor que el valor especificado para alfa, se declara estadísticamente significativo o se dice que está activo. La suma de cuadrados son muy similares a las obtenidas para dos factores; habrá que considerar un subíndice adicional para el tercer factor, esto quedaría de la siguiente manera:

$$SCT = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} Y_{ijkl}^2 - \frac{Y_{....}^2}{N} \tag{3.1}$$

- Donde $N = abcn$ es el total de observaciones en el experimento.

Estructura de ANOVA

$$SC_A = \frac{1}{bcn}\sum_{i=1}^{a} Y_{i\cdots}^2 - \frac{Y_{\cdots\cdot}^2}{N}$$

$$SC_B = \frac{1}{acn}\sum_{j=1}^{b} Y_{\cdot j\cdot\cdot}^2 - \frac{Y_{\cdots\cdot}^2}{N}$$

$$SC_C = \frac{1}{abn}\sum_{k=1}^{c} Y_{\cdot\cdot k\cdot}^2 - \frac{Y_{\cdots\cdot}^2}{N}$$

Estructura de ANOVA

$$SC_{AB} = \frac{1}{cn}\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij\cdot\cdot}^2 - SC_A - SC_B - \frac{Y_{\cdots\cdot}^2}{N}$$

$$SC_{AC} = \frac{1}{bn}\sum_{i=1}^{a}\sum_{k=1}^{c} Y_{i\cdot k\cdot}^2 - SC_A - SC_C - \frac{Y_{\cdots\cdot}^2}{N}$$

$$SC_{BC} = \frac{1}{an}\sum_{j=1}^{b}\sum_{k=1}^{c} Y_{\cdot jk\cdot}^2 - SC_B - SC_C - \frac{Y_{\cdots\cdot}^2}{N}$$

Estructura de ANOVA

$$SC_{ABC} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} Y_{ijk\cdot}^2 - SC_{AB} - SC_{AC} - SC_{BC} - SC_A - SC_B - SC_C - \frac{Y_{\cdots\cdot}^2}{N}$$

$$SC_E = SC_T - SC_A - SC_B - SC_C - SC_{AB} - SC_{AC} - SC_{BC} - SC_{ABC}$$

Estructura de ANOVA

- Los grados de libertad se dan o se definen en el cuadro 5. Una vez realizado el ANOVA, se procede a interpretar los efectos activos y a diagnosticar los supuestos del modelo.

Transformaciones para estabilizar varianza

- En la práctica, algunas variables de respuesta no siguen una distribución normal, sino que se distribuyen, por ejemplo: *Poisson*, *binomial* o *Gamma*, por mencionar estos tres casos. Resulta que en estas distribuciones la media está relacionada con la desviación estándar (variabilidad) y, naturalmente, al cambiar la media de un tratamiento a otro, con ella cambia la variabilidad de respuesta.
- También es cierto que, al suponer normalidad y varianza constante, éstas no se tienen que cumplir de manera estricta, dado que el procedimiento ANOVA es robusto o admite desviaciones moderadas de dichos supuestos.

Transformaciones para estabilizar varianza

- Existen al menos tres formas de solucionar o minimizar el problema por falta de normalidad:
  1. Utilzar métodos de análisis no paramétricos, que no requieren las suposiciones de normalidad y varianza constante.
  2. Hacer el análisis mediante modelos lineales generalizados (GLM), en los que se ajusta un modelo lineal usando otras distribuciones diferentes a la normal, donde la varianza no tiene por qué ser constante.
  3. Hacer el análisis sobre la respuesta transformada a una escala en la que los supuestos se cumplan.

Transformaciones para estabilizar varianza

Cuadro 6: Relación entre la varianza y la media con la transformación apropiada

Relación $\sigma^2 \propto \dots$	**Transformación apropiada**	**Tipo**	**Uso**
$\sigma^2 \propto E(Y)(1-E(Y))$	$Y' = \sin^{-1}(\sqrt{Y})$	Arcoseno	Para proporciones (distribución binomial)
$\sigma^2 \propto E(Y)$	$Y' = \sqrt{Y}$	Raíz cuadrada	Datos tipo Poisson
$\sigma^2 \propto [E(Y)]^2$	$Y' = \ln(Y)$ o $\log_{10}(Y)$	Logaritmo	Varianza proporcional al cuadrado de la media
$\sigma^2 \propto [E(Y)]^3$	$Y' = Y^{-1/2}$	Recíproco raíz cuadrada	—
$\sigma^2 \propto [E(Y)]^4$	$Y' = Y^{-1}$	Recíproco	—

- $\propto$ significa **es proporcional a**.

Sección 5

Diseño factorial general

Diseño factorial general

Hasta el momento para los dos diseños factoriales con 2 y 3 factores puede extenderse facilmente para cuando se tienen más factores. Consideremos los factores A, B, C,...,K, con niveles a, b, c,...,k, respectivamente, donde la letra K detona al f-ésimo o último factor del conjunto a estudiar, no necesariamente el undécimo, que es el lugar de esta letra en el alfabeto.

Con estos niveles y factores se puede construir el diseño factorial general $a \times b \times \cdots \times k$, que consiste de $a \times b \times \cdots \times k$ tratamientos o puntos de prueba.

Con este diseño se puede estudiar f efectos principales, $f\left(f-1\right)/2$ interacciones dobles, $f\left(f-1\right)\left(f-2\right)/\left(3\times 2\right)$ interacciones triples, y así sucesivamente, hasta la única interacción de lo f factores $\left(ABC\cdots K\right)$.

Diseño factorial general

El cálculo del número de interacciones de cierta cantidad m de factores se hace mediande la operación *combinaciones de* f en:

$$\left(\binom{f}{m} = \frac{f\,!}{m\,!\,(f-m)\,!}\right)$$

Que cuenta el número de diferentes maneras de seleccionar m factores de los f, donde $f! = f \times \ (f-1) \times \ \cdots \times 2 \times 1$

En el factor general $a \times b \times \cdots \times k$, se pueden plantear $2^f - 1$ hipótesis que se prueban mediante el análisis de varianza, si se tienen n réplicas.

Diseño factorial general

Cuadro 7: ANOVA para el diseño factorial general $a \times b \times ... \times k$

Fuente de variabilidad	Suma de cuadrados	Grados de libertad
Efecto A	SC_A	$a-1$
$\vdots$	$\vdots$	$\vdots$
Efecto K	SC_K	$k-1$
Efecto AB	SC_{AB}	$(a-1)(b-1)$
$\vdots$	$\vdots$	$\vdots$
Efecto (K-1)K	$SC_{(K-1)K}$	$(l-1)(k-1)$
Efecto ABC	SC_{ABC}	$(a-1)(b-1)(c-1)$
$\vdots$	$\vdots$	$\vdots$
Efecto (K-2)(K-1)K	$SC_{(K-2)(K-1)K}$	$(m-1)(l-1)(k-1)$
$\vdots$	$\vdots$	$\vdots$
Efecto AB $\cdots$ K	$SC_{AB \cdots K}$	$(a-1)(b-1)\cdots(k-1)$
Error	SC_E	$abc\cdots k(n-1)$
Total	SC_T	$(abc\cdots kn)-1$

Diseño factorial general

- La suma de cuadrados totales está dada por:

$$SC_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \cdots \sum_{m=1}^{k} \sum_{r=1}^{n} Y_{ij\cdots mr}^2 - \frac{Y_{***\cdots ***}^2}{N}$$

- La suma de cuadrados de los efectos son:

Diseño factorial general

Sumas de Cuadrados de efectos principales

$$SC_A = \sum_{i=1}^{a} \frac{Y_i^2}{bc \cdots lkn} - \frac{Y_{..}^2}{N}; \quad SC_B = \sum_{j=1}^{b} \frac{Y_j^2}{ac \cdots lkn} - \frac{Y_{..}^2}{N}$$

$$SC_C = \sum_{k=1}^{c} \frac{Y_k^2}{ab \cdots lkn} - \frac{Y_{..}^2}{N}; \quad SC_K = \sum_{m=1}^{k} \frac{Y_m^2}{abc \cdots ln} - \frac{Y_{..}^2}{N}$$

Diseño factorial general

Sumas de Cuadrados de interacciones

$$SC_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij}^2}{c\cdots lkn} - \frac{Y_{..}^2}{N} - SC_A - SC_B$$

$$\vdots$$

$$SC_{(K-1)K} = \sum_{p=1}^{l}\sum_{m=1}^{k}\frac{Y_{pm}^2}{abc\cdots n} - \frac{Y_{..}^2}{N} - SC_{K-1} - SC_K$$

$$\begin{aligned} SC_{ABC} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{Y_{ijk}^2}{d\cdots lkn} - \frac{Y_{..}^2}{N} \\ &\quad - SC_A - SC_B - SC_C - SC_{AB} - SC_{AC} - SC_{BC} \end{aligned}$$

$$\vdots$$

$$\begin{aligned} SC_{AB\cdots K} &= \sum_{i=1}^{a}\cdots\sum_{m=1}^{k}\frac{Y_{ij\cdots m}^2}{n} - \frac{Y_{..}^2}{N} \\ &\quad - SC_A - \cdots - SC_{AB} - \cdots - SC_{AB\cdots(K-1)} \end{aligned}$$

Diseño factorial general

Sumas de Cuadrados del error

$$SC_E = SC_T - SC_A - \cdots - SC_K - SC_{AB} - \cdots - SC_{(K-1)K} - SC_{ABC} - \cdots - SC_{AB\cdots K}$$

- En el ANOVA, para el factorial general $a \times b \times \cdots \times k$ se observa la necesidad de contar con al menos dos réplicas del experimento para calcular la suma de cuadrados del error SC_E y completar toda la tabla ANOVA.
- Sin embargo, esta necesidad de réplicas $(n = 2)$, que se menciona, es para el caso irrel de que interesan $2^f - 1$ efectos.
- Resulta que, con excepción de los diseños con dos o tres factores, en un factorial completo prácticamente nunca interesan todos sus efectos posibles, puesto que en términos generales sólo algunos de ellos están activos.

- El **principio de Pareto**, que en este contexto también se llama prinicpio de **esparcidad de efectos** dice que la mayoría de la variabilidad observada se debe a unos pocos de los efectos posibles; por lo común se deberá a algunos efectos principales e interacciones dobles.

Sección 6

Modelo de efectos aleatorios

Modelo de efectos aleatorios

- Hasta este punto, los modelos de efectos que se han revisado son **modelos de efectos o factores fijos**, lo cual significa que todos los niveles de prueba en cada factor son todos los disponibles para ese factor, o bien, se estudian todos los niveles de interés en ese factor, en ese sentido, los niveles están fijos.
- Con factores dijos, las conclusiones obtenidas sólo son válidas para los niveles de prueba que se estudian en el experimento.

Modelo de efectos aleatorios

- En ocaciones, los niveles de prueba son una muestra aleatoria de la población de niveles posibles. En ese caso, es más apropiado utilizar el modelo de **efectos o factores aleatorios**.
- Un ejemplo de ello, es cuando se prueban cinco instrumentos de medición, pero la población de los mismos es de 100 instrumentos; no es posible experimentar con todos los equipos, entonces se experimenta sólo con cinco de ellos y para obtener conclusiones para toda la pbolación es necesario aplicar métodos apropiados (modelo de efectos aleatorios).

Modelo de efectos aleatorios

- La aplicación de un modelo de efextos aleatorios conlleva la necesidad de considerar la incertidumbre asociado en la elección aleatoria de los niveles de prueba, ya no tiene sentido, para un factor A, preocuparse pro el efecto α_i del nivel i como con los efectos fijos. Lo que ahora tendrá sentido es hablar de la varianza con la que el factor aleatorio contribuye a la variación total; es preciso estimar dicha varianza y probar si su contribución a la variabilidad total es significativa.

El caso de dos factores aleatorios

- Si se consideran dos factores aleatorios A y B de los cuales se prueban a y b niveles elegidos de una población grande de niveles, entonces si los $a \times b$ tratamientos se replican n veces, el modelo de efectos aleatorios es:

Modelo estadístico

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \; i = 1, 2, \cdots, a \; j = 1, 2, \cdots, b \; k = 1, 2, \cdots, n$$

El caso de dos factores aleatorios

- Del modelo estadístico anterior μ es la media general, α_i es el efecto dabido al i-ésimo nivel del factor A, B_j es el efecto del j-ésimo nivel del factor B, $(\alpha\beta)_{ij}$ representa el efecto de interacción de la combinación ij y ϵ_{ijk} es el error aleatorio que, se supone, sigue una distribución normal con media cero y varianza constante $N\left(0, \sigma^2\right)$.
- El modelo es igual al de efectos fijos, pero el hecho de que los efectos sean aleatorios implica que no tiene sentido probar hipótesis directamente sobre tales efectos (medias), sino que ahora el interés se enfoca en estudiar la varianza de dichos efectos.
- Para ello, se supone que los términos α_i, β_i, $(\alpha\beta)_{ij}$ y ϵ_{ijk} son variables aleatorias independientes normales, con media cero y varianzas σ^2_α, σ^2_β $\sigma^2_{\alpha\beta}$ y σ^2 respectivamente.

El caso de dos factores aleatorios

- De esta manera, si se calcular la varianza en ambos lados del modelo, se obtiene el modelo de componentes de varianza dado por:

Modelo de componentes de varianza

$$Var\left(Y_{ijk}\right) = \sigma^2\alpha + \sigma^2\beta + \sigma^2\alpha\beta + \sigma^2$$

- Donde σ_{α}^2, σ_{β}^2 $\sigma_{\alpha\beta}^2$ son las contribuciones de cada efecto a la variación total y se llaman **componentes de varianza**; σ^2 es el componente de varianza debido al error aleatorio.

El caso de dos factores aleatorios

- Las hipótesis son:

Hipótesis

$$H_0 : \sigma^2_\alpha = 0$$
$$H_0 : \sigma^2_\beta = 0$$
$$H_0 : \sigma^2_{\alpha\beta} = 0$$

- Los cálculos necesarios para probar estas hipótesis involucran las mismas sumas de cuadrados del modelo de efectos fijos, la diferencia radica en que para obtener los estadísticos de prueba F_0 aproximados debe tomarse en cuenta que los valores esperados de los cuadrados medios son:

El caso de dos factores aleatorios

Cuadrados medios

$$E\left(CM_A\right) = \sigma^2 + m\sigma^2_{\alpha\beta} + bn\sigma^2_{\alpha}$$
$$E\left(CM_B\right) = \sigma^2 + m\sigma^2_{\alpha\beta} + an\sigma^2_{\beta}$$
$$E\left(CM_AB\right) = \sigma^2 + n\sigma^2_{\alpha\beta}$$
$$E\left(CM_E\right) = \sigma^2$$

El caso de dos factores aleatorios

- De tal forma que para probar las hipótesis, los *estadísticos de prueba apropiados* en el ANOVA son:

Estadísticos de prueba

$$F_0^A = \frac{CM_A}{CM_{AB}}$$
$$F_0^B = \frac{CM_B}{CM_{AB}}$$
$$F_0^{AB} = \frac{CM_AB}{CM_E}$$

El caso de dos factores aleatorios

- Al resolver las ecuaciones dadas por los valores esperados de cuadrados medios para los componentes de varianza, se obtienen estimadores de éstos en función de los cuadrados medios del error, esto es:

Estimadores de σ^2 según CM

$$\hat{\sigma}^2 = CM_E$$

$$\hat{\sigma}^2_{\alpha\beta} = \frac{CM_{AB} - CM_E}{n}$$

$$\hat{\sigma}^2_{\beta} = \frac{CM_B - CM_{AB}}{an}$$

$$\hat{\sigma}^2_{\alpha} = \frac{CM_A - CM_{AB}}{bn}$$