



I.Q.A. LUIS OMAR JAMED BOZA

Dr. Rafael Lucio No. 199-D

Xalapa Eqz., Ver., México C.P. 91000

Tels. (01228) 8155385

lojb33@hotmail.com

lojb@prodigy.net.mx

UNIVERSIDAD VERACRUZANA

**FACULTAD DE INGENIERÍA QUÍMICA
ZONA XALAPA U.V.**

ESPECIALIDAD EN CONTROL DE CALIDAD

ESTUDIO DE POSGRADO

ASIGNATURA

DISEÑO DE EXPERIMENTOS

Edición 1.41

Septiembre 1999 – Agosto 2000

Edición 2.00

Septiembre 2000 – Agosto 2001

Edición 2.10

Septiembre 2001 – Agosto 2002

CONTENIDO

I

| | |
|---|---|
| 1. Introducción | 3 |
| 1.1 Objetivo..... | 4 |
| 1.1.1 Esquema Conceptual | 4 |
| 1.1.2 Aplicación del Diseño de Experimentos..... | 6 |
| 1.1.3 Tipos de Diseños de Experimentos | 7 |
| 1.1.4 Procedimiento General del Análisis de Varianza (ANDEVA) | 8 |
| 1.1.5 Codificación y Escalamiento de datos | 9 |

2

| | |
|--|----|
| 2. Conceptos estadísticos básicos | 10 |
| 2.1 Definición de la Probabilidad..... | 10 |
| 2.2 Siete conceptos en Probabilidad | 12 |
| 2.3 Expectación o Valor esperado..... | 15 |
| 2.4 Algunas distribuciones para la calidad | 16 |
| 2.5 Pruebas de hipótesis..... | 20 |

3

| | |
|--|----|
| 3. Diseño de solo un factor | 25 |
| 3.1 Modelo de efectos fijos | 27 |
| 3.2 Modelo de efectos al azar (aleatorizados)..... | 32 |
| 3.3 Modelos por bloques..... | 32 |
| 3.4 Diseño Cuadrado Latino | 34 |
| 3.5 Estimación de valor faltante | 38 |
| 3.6 Contrastes Ortogonales | 38 |

4

| | |
|--|----|
| 4. Diseños de más de un factor | 43 |
| 4.1 Experimentos factoriales | 44 |
| 4.2 Experimentos 2^k y técnica de Yates..... | 47 |
| 4.3 Experimentos anidados | 52 |
| 4.4 Experimentos Factorial-Anidado..... | 54 |
| 4.5 Regresión Lineal | 56 |
| 4.6 Covarianza y correlación | 59 |

5

| | |
|--|----|
| 5. Contribuciones de Taguchi..... | 64 |
| 5.1 Función de pérdida de Calidad..... | 66 |
| 5.1.1 Determinación de la constante de la función de pérdida | 68 |
| 5.1.2 Estimación de la pérdida esperada..... | 69 |
| 5.2 Resolución de problemas | 70 |
| 5.2.1 Gráficas Lineales | 74 |
| 5.2.2 Análisis Regular | 75 |
| 5.2.3 Relación Señal a Ruido | 77 |
| 5.3 Caso Práctico de Aplicación..... | 78 |
| 5.4 Bibliografía sugerida | 84 |

1. Introducción

Las sociedades en el mundo están en continuos cambios los cuales buscan lograr mejores condiciones y satisfacer las demandas que ellas mismas generan, así también aumenta nuestra consciencia de cohabitar en un sistema cerrado los cuales no permiten hacer derroche de recursos, esto da lugar al pensamiento de Dave Thomas: “No hay límite para la Calidad que se puede producir u otorgar en un servicio, Ni aún en el trabajo más humilde”.

Es un esfuerzo continuo de mejorar, más que un grado fijo de excelencia. Es un resultado. No podemos poseer la Calidad, sólo practicarla. ***La calidad es un estándar de perfección a través del cual juzgamos si llevamos a término lo que nos propusimos, cuando y como dijimos que la haríamos y de manera que satisfaga las necesidades de nuestros Clientes.***

La civilización que actualmente conocemos, desde hace milenios ha fincado su inicial desarrollo en la Agricultura, hemos pasado de un estado de barbarie a ser nómadas recolectores de frutos, posteriormente en el desarrollo de la Agricultura nos convertimos en seres sedentarios y con el paso del tiempo, la Humanidad aprendió a diseñar sus ciclos de preparación, siembra, cuidado y protección y por último la recolección; por lo tanto empezó a seleccionar semillas afín de mejorar sus próximas cosechas.

Este fue el principio de lo que actualmente llamamos Productividad, la infinidad de Factores que intervienen en la Agricultura para obtener resultados óptimos que satisfagan a nuestros Clientes. El estudio de estos factores son los que le dan nacimiento al **DISEÑO DE EXPERIMENTOS**, por un lado el crecimiento de las poblaciones, la exigencia de mejores nutrientes y por otro el desarrollo del Método Científico, coadyuvaron al incremento acelerado de los estudios que promovieron la mejora continua en cantidad y atributos de calidad de los productos del campo.

Actualmente, así como en la Agricultura, el tener una buena cosecha “**DESDE LA PRIMERA VEZ**”, es esencialmente importante para que nuestros productos o servicios puedan adecuadamente entrar en procesos de mejora continua posteriormente, cada vez más existe evidencia que nuestros nuevos procesos deben diseñarse conforme a tres fuerzas que desde la década de los 80's están operando en el Mundo con mayor magnitud, estos son genéricamente: **Cliente, Competencia y Cambio.**

- **LOS CLIENTES ASUMEN EL MANDO:** La fuerza dominante en la relación cliente-servidor, ha definitivamente cambiado a favor del Cliente, hoy los Clientes le dicen a los proveedores qué es lo que quieren, cuando lo quieren y cuanto pagarán.
- **LA COMPETENCIA SE INTENSIFICA:** Ahora no sólo hay más competencia sino que además es de diferentes clases. Los competidores de Nichos han cambiado las bases competitivas, se venden o se dan servicios similares pero unos tienen bases en precio, otros base en otorgamiento de servicio posventa, otros en atributos especiales del producto o servicio.
- **EL CAMBIO SE VUELVE CONSTANTE:** Ya comentamos que los Clientes y la Competencia han cambiado, pero lo mismo ocurre con la Naturaleza misma del cambio, es lo normal y frecuente. Lo importante es que no sólo han disminuido los ciclos rentables de vida de productos y servicios, sino también ha disminuido el tiempo para desarrollar nuevos productos e introducirlos. Hoy las empresas y organizaciones tienen que moverse rápidamente, o no se moverán y desaparecerán.

En este contexto, el Diseño de la Experimentación ha tenido desarrollos espectaculares en diferentes compañías y organizaciones en países como Japón y Estados Unidos, en estos, se han publicado artículos sobre resultados en todas las fases experimentales. Esta asignatura de la Especialidad en Control de Calidad dará en forma breve, una serie de conocimientos que sobre **DISEÑO DE EXPERIMENTOS** se tienen para poder ayudar con los resultados obtenidos y poder facilitar la toma de decisiones.

1.1 Objetivo

Proporcionar a los participantes, los criterios Teóricos y Metodológicos para Diseñar y Analizar en la organización o empresa, acciones y/o experimentos factibles de realizarse cuya finalidad sea el mejoramiento en los Sistemas de Calidad y su comprobación.

Por lo que el objetivo particular que deducimos del anterior es: **ASEGURAR LA MAYOR INFORMACIÓN ÚTIL DEL PROCESO CON EL MENOR COSTO POSIBLE Y LA MENOR INTERRUPCIÓN DEL MISMO** tanto para productos y/o servicios.

1.1.1 Esquema Conceptual

Cuando se carece de información suficiente para resolver un problema, el método de ensayo y error o experimentación es la alternativa para encontrar una solución. A menudo nos enfrentamos con problemas en manufactura, diseño del producto y/o servicio que la única alternativa disponible para resolverlos es la experimentación. Los miembros del equipo que trabajan en el área de Investigación y Desarrollo, o Mejoramiento del Proceso deben encontrar respuestas a preguntas claves como las siguientes:

- ◆ ¿Qué variables influyen en la respuesta del Proceso?
- ◆ ¿Cuál experimento Desarrollar?
- ◆ ¿Qué datos recopilar y cómo interpretarlos?

El conocimiento del Proceso nos ayudará para contestar la primera pregunta, y las dos restantes con la utilización de Diseños Estadísticos Experimentales es factible responderlas, los cuales producen mejores resultados de los que se obtienen al emplear el método de estudiar variable por variable (una variable a la vez), abordado en el familiar estudio de Habilidad o Capacidad del Proceso (Ver asignaturas Control Estadístico de Procesos I y II). Además, son muy eficientes y permiten estudiar distintas variables simultáneamente.

Por otra parte, su aplicación y utilización no es cara y muchas veces se puede desarrollar **"IN SITU"** (en el mismo lugar que se lleva a cabo la producción o ejecución), al menos cuando se compara con otras estrategias de experimentación diseñadas para obtener la misma información, y no implican gastos directos del capital de trabajo.

De hecho, es la aplicación de este tipo de Diseños, lo que ha contribuido en el mejoramiento de la Calidad y Productividad en las Industrias Japonesas, y no la existencia de nuevas instalaciones para manufactura o automatización. Éstas últimas se han obtenido paulatinamente a través de los años mediante la aplicación de Técnicas de Diseños Estadísticos Experimentales y condiciones de rentabilidad.

Una representación esquemática de la acción conjunta del Diseño de Experimentos (DE) y el método científico provoca los siguientes puntos a considerar:

- ◆ Verdadero estado de la naturaleza y comportamiento esperado de las variables de nuestro proceso ya sea en manufactura o en la creación de servicios.
- ◆ Obtención de datos (datos iniciales y nuevos datos), en las condiciones que opera nuestro proceso.
- ◆ Inducción del comportamiento simultáneo de las variables involucradas en el proceso para generar mejores valores de la respuesta, postulando la Hipótesis de funcionamiento.
- ◆ Probar la Hipótesis del rediseño, encontrando valores de las variables que mejoran la respuesta del proceso, tanto en magnitud de las mismas como en reducción de la variación o ruido no controlado.

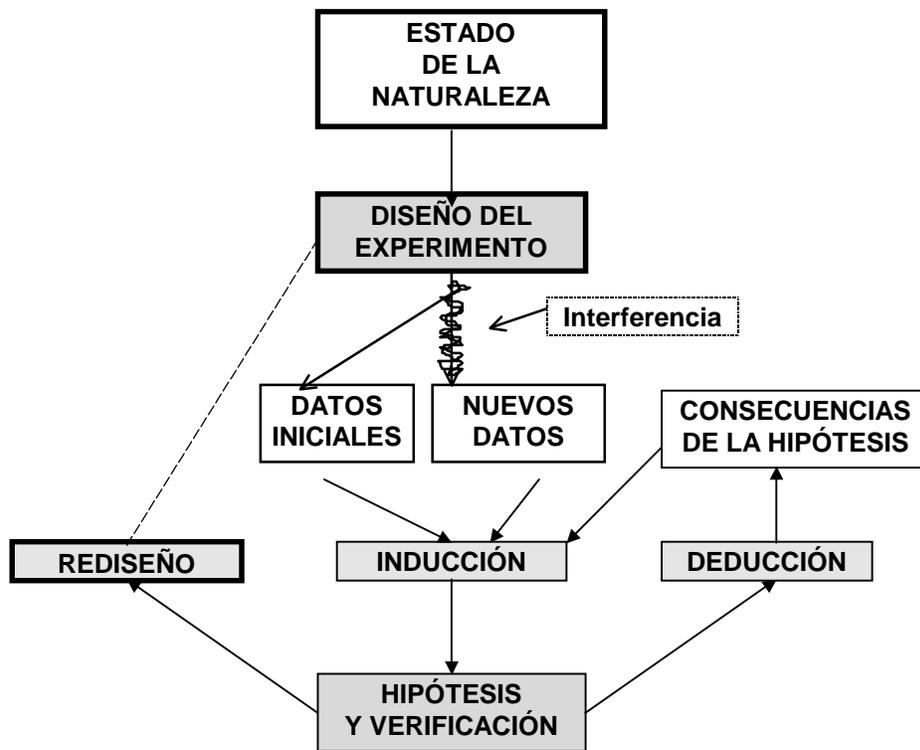


Fig. 1-1

En el esquema anterior el Diseño del Experimento (DE) aparece como si fuera una ventana del estado de la Naturaleza y la zona zigzag se presenta con ruido o interferencia la captación de nuevos datos.

Ya que representa las posibles variaciones producidas tanto por factores no tomados en cuenta en el Diseño o por situaciones de operarios o instrumentos de medición.

1.1.2 Aplicación del Diseño de Experimentos

Como ya comentamos anteriormente, el DE tiene la característica de poder estudiar simultáneamente varios factores a la vez y obtener las posibles interacciones existentes entre ellos así como los efectos provocados de factores que se encuentren anidados (unos dentro de otros), por lo que las aplicaciones del Diseño de Experimentos (DE) implican tener las siguientes características:

CARACTERÍSTICAS IMPORTANTES DEL DISEÑO DE EXPERIMENTOS (DE):

1. Nos permite el estudio de varios factores simultáneos.
2. Estudia las Interacciones entre los factores.
3. Nos permite crear arreglos adecuados de los datos experimentales para el proceso en estudio.
4. Podemos estudiar los efectos producidos por factores que se encuentran dentro de condiciones de otros factores (anidamiento).
5. Nos permite estudiar el efecto resultante de causas disturbantes que se encuentran presentes y con el estudio del proceso lograrlas separar.
6. Nos ayuda a detectar variables ocultas por medio de los efectos que producen.
7. Permite evaluar los efectos de las variables en el desarrollo y el diseño del proceso.
8. Provee un método objetivo de evaluar datos experimentales (y así evitar desviaciones).
9. El experimentador puede diseñar el tamaño del experimento según su Objetivo en el proceso.
10. Introduce el concepto de eficiencia (más información útil con menos trabajo).
11. Provee un método que favorece lograr más confianza para la toma de decisiones en los procesos en estudio.
12. Nos permite tener una aproximación sistemática a las condiciones óptimas de los factores involucrados en el proceso.
13. Simplifica la derivación de funciones matemáticas que representan las relaciones de los factores en el proceso.
14. Genera una formación al personal sin experiencia reafirmando el método científico de atacar los problemas del proceso.
15. Permite atacar los problemas de CENTRALIZACIÓN DE DATOS y los problemas de REDUCCIÓN DE VARIABILIDAD.

El Diseño y Análisis de Experimentos se basa fundamentalmente en la herramienta estadística de el ANÁLISIS DE VARIANZA (ANDEVA o en el lenguaje Inglés ANOVA) que básicamente significa “**EL ANÁLISIS DE LA VARIACIÓN TOTAL ENCONTRADA EN LOS DATOS DEL EXPERIMENTO**”.

Bajo la característica 15, el DE y el ANDEVA nos permiten tener dos enfoques para atacar los problemas en el área de CALIDAD, estos son:

- I. Detectar corrimiento o sesgo debido al efecto (nivel) de un cierto factor en estudio, con respecto a la media de todos los datos.
- II. Al detectar el comportamiento de los efectos (niveles) provocados por los factores en estudio y conociendo el proceso, entonces podemos ajustarlo en condiciones de mayor consistencia (menor variabilidad del proceso, implica proceso robusto).

NOTA: Por ROBUSTO o CONSISTENTE debemos entender que el producto y/o servicio o en su caso el proceso, funciona de manera a como se planeó, y es relativamente insensible a factores difíciles de controlar (Taguchi & Wu, 1980 y Taguchi 1986).

En la estructuración del Diseño y Análisis de Experimentos (DAE) para algún proceso de manufactura y/o servicios al Cliente, es necesario tener presente los siguientes puntos:

1. **PLANEACIÓN DEL EXPERIMENTO:** En primer lugar debemos tener definido el PROBLEMA Y EL OBJETIVO a lograr, en segundo lugar por conocimiento de la situación actual del proceso debemos establecer los DATOS y la forma de obtenerlos que se requiere para resolver el problema y obtener el objetivo.
2. **DISEÑO DEL EXPERIMENTO:** Conforme al conocimiento del proceso y habiendo definido la VARIABLE DE RESPUESTA que analizaremos, entonces seleccionar el tipo de Diseño adecuado conforme al problema a resolver y objetivo a lograr, identificando los factores involucrados y sus relaciones (de independencia y condicionalidades).
3. **ANÁLISIS DEL EXPERIMENTO:** La interpretación de los resultados deberá ser conforme a las condiciones elaboradas en los dos puntos anteriores. Dicha interpretación podrá hacerse vía Gráficas de Control, o pruebas de Hipótesis sencillas si SOLO UNA VARIABLE SE ENCUENTRA EN ESTUDIO, o si existen más de una en estudio, entonces podrá hacerse vía ANÁLISIS DE VARIANZA.

1.1.3 Tipos de Diseños de Experimentos

La selección del tipo de Diseño de Experimento a efectuar obviamente depende en la práctica del CONOCIMIENTO Y FAMILIARIZACIÓN que se tenga del Proceso en estudio y al Objetivo a lograr, a continuación presentaremos sin pretender ser exhaustivo los tipos más comunes y algunas características estudiadas en ellos:

| TIPOS DE EXPERIMENTOS | CARACTERÍSTICAS ESTUDIADAS |
|---|---|
| 1) De Comparación Simple | a) Tendencia Central del Proceso b) Dispersión del Proceso |
| 2) De un solo Factor | a) Niveles (o efectos del factor) principales con efectos fijos o efectos al azar. |
| 3) Diseños Factoriales (más de un factor) | a) Niveles principales de los Factores b) Todas las Interacciones entre Factores |
| 4) Diseños Jerárquicos o Anidados | a) Efectos principales y condicionalidades |
| 5) Diseño Bloque Completo al azar Cuadrado Latino Cuadrado Greco-Latino | a) Efectos del Factor principal sin Interac. b) Eliminación de fuentes de variabilidad ajenas al estudio |
| 6) Diseño Factorial Anidado | a) Efectos principales y condicionalidades b) Algunas Interacciones |
| 7) Diseños Factoriales Fraccionarios | a) Efectos principales b) Algunas Interacciones |

1.1.4 Procedimiento General del Análisis de Varianza (ANDEVA)

Como ya comentamos anteriormente (sección 1.1.2), el ANDEVA significa el Estudio y Análisis de la Variación Total encontrada en los datos del Experimento.

Bajo este significado el ANDEVA desglosa las diferentes fuentes de variación consideradas en el modelo del Diseño de Experimentos y estima las variaciones de cada una de ellas, haciendo prueba de hipótesis en cada uno de los factores y/o interacciones entre ellos.

El procedimiento genérico utilizado en el ANDEVA es el siguiente:

1. La variación total encontrada en los datos, la desglosamos en los factores e interacciones en estudio y llegamos a expresiones tales como: $SS_T = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_E$ en donde las letras SS quieren decir Suma de Cuadrados (del inglés Squared Sum) y los subíndices corresponden a los factores e interacciones en estudio A, B, C, AB, AC, BC, ABC y E.
2. Obtenemos los grados de libertad de cada factor e interacción correspondiente.
3. Dividimos los resultados obtenidos del punto 1 entre los grados de libertad obtenidos en el punto 2, y a esta división se le conoce con el nombre de cuadrática media o simplemente cuadrado medio $CM = \frac{SS}{G.L.}$ de los factores e interacciones en estudio.
4. Ya obtenidos en el punto 3 los cuadrados medios de los factores e interacciones en estudio, habrá notado estimado lector que en el punto 1 seguimos en orden alfabético acostumbrado, exceptuando la letra E que está representando las variaciones no justificadas por el desglose de factores A, B y C que se encuentran en estudio y sus interacciones AB, AC, BC (de dos factores) y la ABC (de los tres factores) a este término se le llama genéricamente ERROR.
5. En la mayoría de los casos, exceptuando en situaciones que los niveles de los factores sean obtenidos al azar y obteniendo sus valores esperados correspondientes, los cuadrados medios de los factores e interacciones se dividirán entre el cuadrado medio del error, ya que éste último estima el valor de σ^2 de la población total de datos.
6. La división del cuadrado medio del factor entre el cuadrado medio del error es en si mismo una división de varianzas, por lo tanto representa el estadístico F_0 .
7. La prueba de hipótesis que corresponda por el modelo utilizado se probará con el estadístico obtenido del punto 6, es decir, el F_0 contra las tablas de distribución de Fisher a un determinado nivel de significación α (error tipo I). Si resultan ser significativo los valores, entonces se concluye que los niveles del factor en estudio SI MODIFICAN LOS VALORES DE LA VARIABLE RESPUESTA QUE SE ESTÁ MANEJANDO.

Obviamente este procedimiento de 7 puntos quedará más claro en las siguientes unidades las cuales haremos ejercicios y reafirmaremos las decisiones que deben tomarse en cada modelo de Diseño de Experimentos.

1.1.5 Codificación y Escalamiento de datos

En el ANDEVA, frecuentemente tenemos valores de la variable de respuesta de algún Diseño de Experimentos, los cuales son difíciles de manipular, ya sea por ser de gran magnitud o por valores demasiado pequeños, y fácilmente podríamos cometer errores numéricos, ya sea por redondeos o por mala introducción de los datos en las calculadoras, para estos casos es factible poder hacer; ya sea una Codificación de los datos, o bien, un Escalamiento de ellos.

El ANDEVA trabaja con sumas de cuadrados, que tiene la siguiente expresión en el caso de un factor, pero es válido la codificación y escalamiento para más factores:

$$\mathbf{SS}_T = \mathbf{SS}_A + \mathbf{SS}_E \quad \text{Ec. 1-1}$$

$$\mathbf{SS}_T = \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{i,j} - \bar{\bar{x}})^2 \quad \text{Ec. 1-2}$$

$$\mathbf{SS}_A = \sum_{i=1}^a n_i (\bar{x}_i - \bar{\bar{x}})^2 \quad \text{Ec. 1-3}$$

$$\mathbf{SS}_E = \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2 \quad \text{Ec. 1-4}$$

La fuente de variación total \mathbf{SS}_T se desglosa en dos fuentes de variación, una justifica la variación de los (**a**) niveles del factor A y lo no justificado por el modelo, que genéricamente le llamamos error, en la ecuación 1-1: Dándonos como resultado del desglose las expresiones 1-2, 1-3 y 1-4.

La codificación implica el sumar o restar un número constante, digamos C, a todos los datos $X_{i,j}$ experimentales, éste se reflejará tanto en los datos originales como en la media de cada nivel (i) y en la gran media (i,j), por lo que las expresiones de Sumas de Cuadrados tendrán los mismos valores.

El escalamiento implica el multiplicar o dividir un número constante, digamos C_2 , a todos los datos $X_{i,j}$ experimentales, esto creará un coeficiente constante ya sea C^2 o $1/C^2$ que se eliminará cuando dividamos el cuadrado medio del factor A entre el cuadrado medio del error. El utilizar estas técnicas de codificación y escalamientos, evita cometer errores en los cálculos.

NOTA: El expositor, desarrollará en clase la metodología completa y ejemplos de utilización de codificación y escalamiento de datos en el ANDEVA.

2. Conceptos estadísticos básicos

En esta unidad, en forma breve repasaremos conceptos básicos de Probabilidad y Estadística, estos son:

- **Definición de la Probabilidad:** En él, se presenta la idea de proporción de algo favorable entre el total de elementos que pueden suceder.
- **Siete conceptos fundamentales en Probabilidad:** Los tres primeros conceptos definen los valores numéricos que tiene la probabilidad. Los siguientes cuatro conceptos son operativos y nos permiten resolver problemas.
- **Concepto de Expectación o Valor Esperado:** Establece los parámetros característicos que definen a alguna distribución de probabilidad, ya sea discreta o continua.
- **Algunas distribuciones de Probabilidad con interés para la Calidad:** Estas que repasaremos brevemente son: La Binomial, La Hipergeométrica, La Poisson (SON DISCRETAS); y La Normal, La Exponencial, La Gamma (SON CONTINUAS).
- **Inferencia estadística y Pruebas de Hipótesis:** En esta parte se plantean las pruebas de hipótesis de medias y de varianzas; las distribuciones t (the student), X^2 (ji-cuadrada) y F (de Fisher).

2.1 Definición de la Probabilidad

La conceptualización de la probabilidad, es la idea de proporción y refleja una división, en donde el numerador representa el número de elementos favorables al evento que deseamos obtener su probabilidad, entre el divisor que representa el número de elementos totales que existen en la población o universo de datos. Para llegar a esta división, tenemos tres caminos u opciones, las cuales deben ser consistentes con la realidad de los comportamientos en los datos. Estos caminos u opciones son:

1. **Heurístico:** Básicamente está relacionado con la experiencia del investigador y es intuitivo.
2. **Formal:** Se encuentra basado en la posibilidad de crear un cierto modelo, ya sea físico o mental que no necesariamente se apoya en la experimentación.
3. **Frecuencial:** Este se encuentra basado mediante largos procesos de experimentación y la idea de probabilidad se relaciona con la frecuencia relativa de ocurrencia del evento en estudio.

Estos tres caminos conllevan a la probabilidad, en su resultado final, deben coincidir con la realidad de los datos y obtendremos un valor numérico. Ver la siguiente figura 2-1.

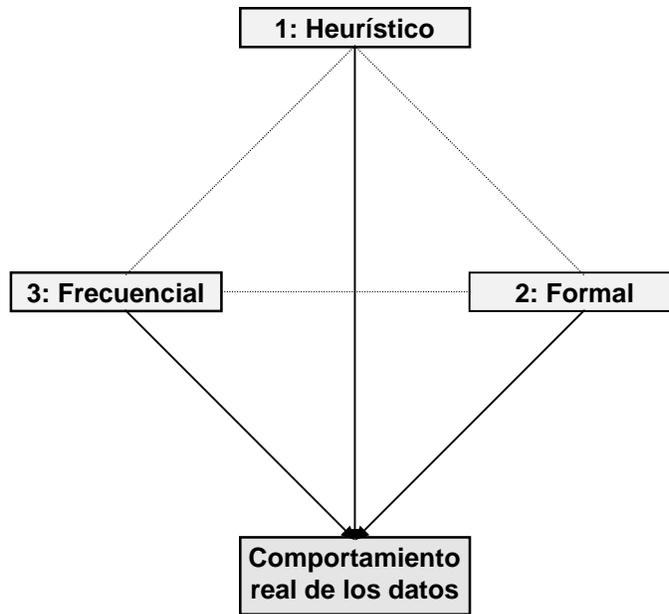


Fig. 2-1

La percepción heurística de un fenómeno, puede darnos intuitivamente la proporción en que estimemos el evento en estudio, es decir, un posible porcentaje de las cosas; pero no es suficiente esta percepción para validar numéricamente la probabilidad del evento.

Si a la percepción anterior la formalizamos creando un modelo que cumpla con la experiencia intuitiva, empezamos a respaldar la posible proporción numérica que encontremos, pero todavía no es suficiente para validar y verificar el resultado, sólo utilizando conjuntamente las opciones 2 y 3 se nos permite con la experimentación lograr confirmar el modelo utilizado y poder validar, verificando el resultado numérico que encontremos de la probabilidad.

Si (x) es el evento definido en estudio de la población, $n(x)$ es el número de elementos o veces favorables al evento (x) y (m) es el número total de elementos o veces que se lleva a cabo el experimento, entonces la Probabilidad está dada por la siguiente igualdad.

$$P(x) = \frac{\text{Fav o rables}}{\text{To ta les}} = \frac{n(x)}{m} = \lim_{m \rightarrow \infty} \frac{n(x)}{m}$$

2.2 Siete conceptos en Probabilidad

Como ya comentamos anteriormente (al inicio de la unidad 2), los tres primeros conceptos definen los valores numéricos que pueden tenerse cuando manejamos la probabilidad, ya habiendo definido el universo o población y el evento en estudio, estos son:

1. Define el menor valor posible de la probabilidad y se relaciona con el conjunto vacío.

$$P(\phi) = 0.00$$

La ausencia de elementos favorables al evento x , define el conjunto vacío e implica el valor numérico de la probabilidad de cero.

2. Define el mayor valor posible de la probabilidad y se relaciona con el universo o población de datos completa.

$$P(U) = 1.00$$

La presencia completa de elementos favorables al evento x , define a la población e implica el valor numérico de uno.

3. Define el intervalo de valores que podrá tomar la probabilidad del evento x .

$$0.00 < P(x) < 1.00$$

El valor de la probabilidad del evento x , deberá de encontrarse siendo mayor que cero y a su vez menor que uno.

Los siguientes cuatro conceptos son operativos y nos permiten manejar los valores, obteniendo resultados y su interpretación práctica de lo que representan. Estos también se les conoce con los siguientes nombres: Principio Aditivo, Principio Multiplicativo, Regla Condicional y el último que es el Teorema de Bayes.

Si asociamos las áreas de eventos A y B en un diagrama de Venn, estos eventos pueden ser mutuamente excluyentes o no, pero el valor de su probabilidad queda definido en la siguiente forma:

$$P(A) = \frac{\text{Área}(A)}{\text{Área}(U)} \quad \text{y} \quad P(B) = \frac{\text{Área}(B)}{\text{Área}(U)}$$

Para eventos en general A y B la probabilidad de la unión de ellos es la siguiente:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Ya que el área de la intersección se duplica debe restarse la probabilidad de la intersección una vez, para eventos mutuamente excluyentes, la probabilidad de la intersección es cero ya que es conjunto vacío.



Fig. 2-2

En la figura 2-2 de la izquierda los eventos no son mutuamente excluyentes por lo que la probabilidad de la unión se obtiene con la fórmula general anterior, la figura de la derecha muestra los eventos A y B mutuamente excluyentes, por lo que la probabilidad de la intersección tiene valor de cero.

4. **Principio Aditivo:** Se establece para eventos mutuamente excluyentes, por lo que los valores de la probabilidad de intersecciones valen cero, la fórmula de unión de eventos es la siguiente:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

Ser mutuamente excluyente los eventos A y B, implica que si sucede A es imposible que suceda B simultáneamente y viceversa.

5. **Principio Multiplicativo:** Se establece para eventos independientes entre si, e implica la simultaneidad de ocurrencia de eventos, cuando 2 o más eventos suceden simultáneamente y son independientes entre si, la fórmula de intersecciones de eventos es la siguiente:

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$$

Ser independiente entre si los eventos A y B, implica que el suceder A no modifica la probabilidad de que suceda B, y viceversa.

6. **Probabilidad Condicional:** Cuando un cierto evento A modifica su probabilidad por el suceso del evento B, entonces la probabilidad de A está condicionada al evento B. Acostumbramos a describir este hecho en la siguiente forma: $P(A/B)$ la probabilidad de A dado que ya sucedió B; cuando el evento que sucede es A y se modifica la probabilidad de B, lo describimos en la siguiente forma: $P(B/A)$ la probabilidad de B dado que ya sucedió A.

$$P(A / B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B / A) = \frac{P(A \cap B)}{P(A)}$$

O bien en su forma de multiplicación la cual da la probabilidad de la intersección de los eventos A y B.

$$P(A \cap B) = P(A) \times P(B / A) = P(B) \times P(A / B)$$

La diferencia sustancial entre el principio multiplicativo y la probabilidad condicional, es primeramente que el concepto 5 (principio multiplicativo) está definido para eventos que son independientes entre si; en cambio el concepto 6 (probabilidad condicional), es para eventos A y B en general.

Ya que si aplicamos la independencia de eventos en el concepto 6, obtenemos si A y B son independientes entre si lo siguiente:

$$P(A / B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

La expresión anterior demuestra que el valor de la probabilidad del evento A no es modificada si B sucede o deja de suceder, que es lo que implica la independencia.

Como corolario de lo anterior, podemos darnos cuenta que lo general es la condicionalidad y un caso particular sólo cuando se supone el concepto 5 para que cumpla la independencia se deberá de cumplir: **P(A intersección B) = P(A) x P(B)**.

7. **Teorema de Bayes:** Es la generalización de la probabilidad condicional, para plantear el teorema de Bayes, necesitamos definir primeramente lo que significa una partición:

Una partición, es una división completa del universo o población (como las rebanadas de un pastel), en la cual todos los eventos son mutuamente excluyentes pero su unión forman a la población, si k eventos B forman una partición y A es un evento definido en ese universo, en numerosas situaciones prácticas no puede ser computada la probabilidad de A directamente, entonces nos ayudamos con los eventos B en la siguiente manera para calcular la Probabilidad de A.

$$P(A) = P(B_1)P(A / B_1) + P(B_2)P(A / B_2) + \dots + P(B_k)P(A / B_k)$$

Cada término sumado, representa la intersección del evento A con cada uno de los B's de la partición, ver la siguiente figura (A.I.B , representa A intersección B):

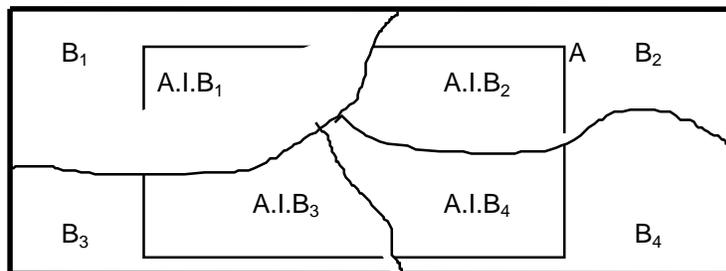


Fig. 2-3

Si Br es un evento fijo de la partición B y se desea obtener la P(Br/A), entonces el teorema de bayes plantea lo siguiente:

$$P(Br / A) = \frac{P(Br \cap A)}{P(A)} = \frac{P(Br) \times P(A / Br)}{\sum_{i=1}^k P(B_i) \times P(A / B_i)}$$

NOTA: El expositor generará ejemplos y utilidades en el área de calidad de lo expuesto en esta sección.

2.3 Expectación o Valor esperado

Es importante para la Calidad, recordar el concepto del valor esperado o expectativa, desde la idea formal y frecuencial de la probabilidad.

Desde el punto de vista frecuencial, el valor esperado es el resultado que esperamos obtener si efectuamos muchas veces el experimento con las mismas condiciones y en el mismo ambiente de experimentación.

Desde el punto de vista formal, el valor esperado nos define los parámetros característicos del comportamiento de los datos en la población, si (x) es una variable aleatoria, $f(x)$ es la función de densidad de probabilidad en distribuciones continuas y $P(x)$ es el valor de la probabilidad del evento (x) en distribuciones discretas, siendo $H(x)$ una función de la variable aleatoria (x) , entonces el valor esperado tiene las siguientes expresiones:

$$E(H(x)) = \int_{-\infty}^{\infty} H(x)f(x)dx \quad \text{para continuas}$$

$$E(H(x)) = \sum_U H(x)P(x) \quad \text{para discretas}$$

Para las áreas de Calidad, los valores esperados más importantes son los que nos definen la media y varianza de la población que son cuando la $H(x)$ tiene los siguientes valores:

$$\text{Si } H(x) = x \text{ entonces } E(x) = \mu$$

$$\text{Si } H(x) = (x - \mu)^2 \text{ entonces } E((x - \mu)^2) = V(x) = E(x^2) - [E(x)]^2 = \sigma^2$$

El término de varianza es utilizado extensivamente en las áreas de calidad, por lo que es importante conocer al operador **V** el cual se define en términos del valor esperado u operador **E**.

2.4 Algunas distribuciones para la calidad

En esta sección repasaremos brevemente, la Binomial, la Hipergeométrica, la Poisson, la Normal, la Exponencial, y la Gamma, las cuales en forma de tabla describiremos parámetros y expresiones:

| Distribución | Parámetros | Media = E(x) | Varianza = V(x) |
|-----------------|---|--------------------------|--|
| Binomial | n = 1,2,3, ... p = prob. de éxito | np | np(1-p) |
| Hipergeométrica | N = 1,2,3, ... n = 1,2,3, ... , N D = 1,2,3, ... , N | $n \frac{D}{N}$ | $n \frac{D}{N} \left[1 - \frac{D}{N} \right] \frac{N-n}{N-1}$ |
| Poisson | $\alpha = np$ o bien $\alpha = \lambda x$ o bien $\alpha = \lambda t$ | α | α |
| Normal | $\mu =$ media $\sigma^2 =$ varianza | μ | σ^2 |
| Exponencial | $\lambda =$ tasa promedio $\lambda =$ discret./contin. | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Gamma | $\alpha =$ par. de forma $\lambda =$ par. de escala | $\frac{\alpha}{\lambda}$ | $\frac{\alpha}{\lambda^2}$ |

La **distribución Binomial** tiene las siguientes características:

- En cada experimento sólo tiene dos eventos mutuamente excluyentes que pueden suceder.
- Los experimentos entre si, son independientes, (la probabilidad de éxito no cambia de valor de un experimento a otro).
- El tamaño del universo o población **N**, es grande con reemplazo o infinito.

Su expresión es la siguiente:

$$P(x) = B(x,n,p) = C_x^n p^x (1-p)^{n-x}$$

La **distribución Hipergeométrica** tiene las siguientes características:

- En cada experimento sólo tiene dos eventos mutuamente excluyentes que pueden suceder.
- Cada experimento hipergeométrico depende del resultado del experimento anterior, (la probabilidad de éxito se modifica de un experimento a otro).
- El tamaño del universo o población **N**, es pequeño sin reemplazo.

Su expresión es la siguiente:

$$P(x) = H(x,n,D,N) = \frac{C_x^D \times C_{n-x}^{N-D}}{C_n^N}$$

La **distribución Poisson** es discreta ya que su variable aleatoria (x) lo es, su parámetro α , representa el número promedio de eventos discretos.

Ésta distribución, es muy útil en las áreas de Calidad ya que refleja el comportamiento de procesos secuenciales y calcula la probabilidad de que suceda un cierto número de eventos (x) manteniendo cualquier otro factor continuo constante o fijo, se comprende mejor colocando ejemplos de utilización:

- **Llegadas de lotes de Producción en cada hora de manufactura:** La variable lotes de producción es discreta (x), pero están fijas al tiempo continuo de una hora.
- **Llegadas de órdenes de pedido para ser procesados en un día:** La (x) discreta son órdenes de pedido, pero están fijadas por día.
- **Llegadas de clientes al banco para efectuar operaciones financieras por hora:** El número de clientes que llegan al banco es la variable discreta (x) y el tiempo está fijo.
- **No. de clientes al cine para efectuar la compra del boleto por cada 15 minutos:** Las personas que llegan a la ventanilla es la variable discreta (x) y el tiempo de 15 minutos está fijo.

El parámetro alfa (α), de Poisson tiene unidades de eventos discretos y para los ejemplos de utilización anteriores tiene la siguiente expresión:

$$\alpha = \lambda t$$

Para el caso de aproximación a la Poisson proviniendo de la Binomial, el parámetro alfa tiene la expresión de $\alpha = np$.

La (t) representa tiempo fijo, y la (λ) representa la tasa promedio de eventos discretos por unidad de tiempo o evento continuo fijo.

Su expresión es la siguiente:

$$P(x) = \frac{e^{-\alpha} \alpha^x}{x!}$$

En las distribuciones de probabilidad que la variable aleatoria (x o t) sea continua, la probabilidad de un intervalo desde un valor (a) hasta un valor (b) viene dado por el área bajo la curva de la función de densidad de probabilidad, si $f(x)$ es dicha función, entonces la probabilidad del intervalo es la siguiente:

$$P(a \leq x \leq b) = \int_a^b f(x) dx$$

Las distribuciones Normal, Exponencial y Gamma tienen las siguientes expresiones:

Normal, con media μ y desviación estandar σ .

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Exponencial, con media $1/\lambda$ y varianza $1/\lambda^2$.

$$f(t) = \lambda e^{-\lambda t}$$

Gamma, con media α/λ y varianza α/λ^2 .

$$f(t) = \frac{\lambda}{\Gamma(\alpha)} (\lambda t)^{(\alpha-1)} e^{-\lambda t}$$

La distribución Poisson calcula la probabilidad de que un cierto número (x) de fallas sucedan en un tiempo determinado (fijo), en cambio la Exponencial calcula la probabilidad de que en un cierto intervalo de tiempo (variable t) suceda exactamente la falla, es decir la Poisson calcula FALLAS y la Exponencial calcula intervalo de TIEMPO para una falla.

Se encuentran relacionadas la Gamma con la Exponencial ya que una suma de un número (α) de exponenciales con un mismo valor (λ) generan una distribución Gamma. Un ejemplo de lo anterior se encuentra en el siguiente ejercicio.

- Se tiene un sistema que opera redundantemente con tres unidades A,B y C, éste sistema tiene un switch que conecta el punto 1 con el punto 2, las unidades provienen de un lote el cual tiene una tasa promedio de falla de $\lambda=1/100$ (una falla en cada 100 hrs. de funcionamiento) es decir $\lambda=0.01$ fallas/hora, el switch opera inmediatamente al suceder la falla, su posición se encuentra para la unidad A, sólo al fallar ésta, inicia su funcionamiento la unidad B y sólo al fallar el switch conecta a la unidad C. La condición mínima de calidad que se le solicita antes de dar un descanso al sistema es que opere 12 horas continuas.
- a) Con la distribución Poisson calcule la probabilidad de que la unidad tenga x=0 fallas en 12 horas (fijas) de tiempo de funcionamiento (Fiabilidad individual de cada unidad del sistema).

Como las tres unidades son del mismo lote y tienen una misma λ (0.01 falla/hrs), entonces el alfa individual de cada unidad y la prob. de cero fallas es:

$$\alpha = \lambda t \text{ (de Poisson)} = (.01)(12) = .12$$

$$P(x = 0) = \frac{e^{-.12} (.12)^0}{0!} = .886920436 \approx .8869204$$

El resultado anterior quiere decir que esperamos de cada 1000 veces que cada unidad trabaje 12 horas continuas, 887 veces aprox. será fiable que funcione la unidad y por complemento esperamos que falle la unidad 113 aprox. de las 1000.

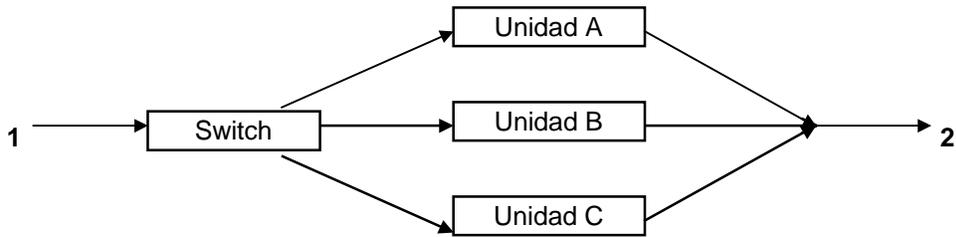


Fig. 2-4

- b) Con la distribución Exponencial calcule la probabilidad de que el tiempo cuando ocurra la falla sea mayor de 12 horas (fiabilidad para cada unidad del sistema).

$$P(t > 12) = \int_{12}^{\infty} \lambda e^{-\lambda t} dt = [e^{-\lambda t}]_{\infty}^{12} = .886920436 \approx .8869204$$

Si cada unidad es independiente entre si y tenemos una variable (t) que es la vida del sistema, puede ser representada como la suma de las vidas de los subsistemas: $t = t_1 + t_2 + t_3$, en donde la densidad de cada subsistema tiene la siguiente expresión:

$$g(t_i) = \lambda e^{-\lambda t_i}$$

La variable (t) tiene una distribución Gamma con alfa igual a 3 y lambda igual a 0.01, la expresión de la densidad de probabilidad para este ejemplo es:

$$f(t) = \frac{0.01}{2!} (0.01t)^2 e^{-0.01t} \quad t > 0$$

La probabilidad de que el sistema funcionará al menos (t) horas es la función de Fiabilidad, que para este ejemplo es:

$$R(t) = 1 - F(t) = \sum_{k=0}^2 e^{-0.01t} \left(\frac{(0.01t)^k}{k!} \right) \quad t > 0$$

- c) Calcular la fiabilidad del sistema A, B y C para un tiempo de 12 horas.

$$R(t) = e^{-0.01t} \left[1 + (0.01t) + \frac{(0.01t)^2}{2} \right] = 0.9997367$$

2.5 Pruebas de hipótesis

Una hipótesis estadística es una afirmación acerca de los parámetros de una población, los cuales deben verificarse con una prueba, la hipótesis que se desea probar comúnmente la llamamos hipótesis nula, lo no contenido en la hipótesis se localizará en la hipótesis alternativa.

Generalmente identificamos la hipótesis nula como H_0 y la hipótesis alternativa como H_1 , en la toma de decisiones de la prueba de hipótesis, podemos cometer dos tipos de error.

Ocurre un error de tipo I, cuando la hipótesis nula es rechazada siendo verdadera. Si la hipótesis nula no es rechazada cuando es falsa se comete un error de tipo II. La probabilidad de estos errores se conocen como alfa y beta.

$$\alpha = P(\text{error tipo I}) = P(\text{rechazar } H_0/H_0 \text{ es verdadera})$$

$$\beta = P(\text{error tipo II}) = P(\text{no rechazar } H_0/H_0 \text{ es falsa})$$

Es muy conveniente recordar también la expresión del poder de la prueba de hipótesis, que está basada en el valor de beta.

$$\text{Poder} = 1 - \beta = P(\text{rechazar } H_0/H_0 \text{ es falsa})$$

Las pruebas de Hipótesis más comunes prueban MEDIAS (con varianza conocida o desconocida), y también prueban VARIANZAS (contra un valor dado o relación de varianzas de dos poblaciones que se están comparando).

A) PRUEBA DE HIPÓTESIS DE MEDIAS:

A.1) **Con Varianza conocida**, se utiliza el estadístico z_0 , y se compara contra valores de la Normal Estándar.

A.2) **Con Varianza desconocida**, se utiliza el estadístico t_0 , y se compara contra los valores de la Distribución del estudiante (The Student).

A.3) **Por pares de medias**, se utiliza el estadístico z_0 y se compara contra los valores de la Normal (si es varianza conocida), o contra la t_0 (si es varianza desconocida).

B) PRUEBAS DE HIPÓTESIS DE VARIANZAS:

B.1) **Contra un valor dado de varianza**, se utiliza el estadístico X_0^2 , y se compara contra los valores de la Distribución Ji-Cuadrada.

B.2) **De relación de varianzas**, se utiliza el estadístico F_0 , y se compara contra los valores de la Distribución de Fisher.

Dentro de los conceptos de Inferencia Estadística, podemos demostrar que cuando tomamos datos de una población, y analizamos dichos datos, el comportamiento de ellos tiende a ser Normal (estudio de los experimentos de Shewhart, Control de Procesos I). Aquí nos interesa recordar la relación que existe entre la Población original de datos (UDDO) y la Población que se genera a partir de la anterior que es la Población de muestras posibles (UDMP), que es en el que realmente trabajamos cuando hacemos una prueba de hipótesis.

Si μ y σ son la media y desviación estándar de la Población original de datos (UDDO) y μ_x , σ_x son la media y desviación de la Población de muestras posibles (UDMP) entonces la relación entre ambas Poblaciones es:

$$\mu_x = \mu \quad y \quad \sigma_x^2 = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

NOTA: El expositor desarrollará ejemplo ilustrativo de esta relación de poblaciones.

En donde la **N** es el tamaño de la Población original de datos y **n** es el tamaño de la muestra que se haya seleccionado, el primer término de la varianza tiende a valer 1 cuando **N** es muy grande o infinito como en el caso de las distribuciones continuas.

Dicho término $(N-n)/(N-1)$ se le conoce con el nombre de factor de corrección por reemplazo. Los estadísticos para las pruebas de hipótesis para medias son:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \quad y \quad t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

En una prueba de hipótesis, el valor del error tipo I, se fija para tomar la decisión (rechazar o no-rechazar), y se acostumbra llamar nivel de significación de la prueba (α). El valor del otro posible error que podemos cometer en nuestra decisión es el beta (β) y éste se puede tener bajo control sólo desde el diseño de la misma prueba de hipótesis, por medio del manejo de la curva característica de operación y la condición de calidad que se requiera cumplir, ya sea que postulemos un valor máximo del error beta para una diferencia de lo que vamos a considerar verdadero con respecto a lo que vamos a considerar falso en la prueba; entonces como resultado determinaremos el tamaño de muestra que se requiere para que cumpla el criterio de calidad establecido.

O bien, podemos obtener el valor de beta (β), teniendo un tamaño de muestra (**n**) y la condición a cumplir de calidad es la diferencia entre lo verdadero y falso a considerar.

NOTA: Después de haber hecho ejemplos de pruebas de hipótesis el expositor ilustrará y calculará la forma de controlar éste segundo tipo de error que podemos cometer.

A continuación, presentaremos en forma de tabla las pruebas de hipótesis, recordándoles que la decisión de rechazar se rige por el planteamiento que se haya colocado en la hipótesis alternativa (ya que se rechazó la nula). Si no logramos rechazar entonces podemos considerar que el planteamiento colocado en la hipótesis nula se está cumpliendo.

Tabla 2-1 Pruebas de medias, con varianza conocida.

| HIPÓTESIS | ESTADÍSTICO DE PRUEBA | CRITERIO P/RECHAZO |
|---|--|------------------------|
| $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ | $Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ | $ Z_0 > Z_{\alpha/2}$ |
| $H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$ | | $Z_0 < -Z_\alpha$ |
| $H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$ | | $Z_0 > Z_\alpha$ |
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ | $Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ | $ Z_0 > Z_{\alpha/2}$ |
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$ | | $Z_0 < -Z_\alpha$ |
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$ | | $Z_0 > Z_\alpha$ |

Estas pruebas de medias son útiles para cuando estamos diseñando el experimento, ya que generalmente en la práctica no conocemos la varianza de la población en estudio, pero si tenemos experiencia en el proceso, es posible utilizarlas adecuadamente.

NOTA: El expositor, generará ejemplos ilustrativos en clase de este tipo de prueba.

Para las pruebas con varianza desconocida, en donde se comparan dos poblaciones (1 y 2), se puede considerar una “**varianza combinada**” en el estadístico (t) sólo si podemos suponer que son iguales las varianzas poblacionales.

Tabla 2-2 Pruebas de medias, con varianza desconocida.

| HIPÓTESIS | ESTADÍSTICO DE PRUEBA | CRITERIO P/RECHAZO |
|---|---|---------------------------|
| $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ | $v = n - 1$ | $ t_0 > t_{\alpha/2, v}$ |
| $H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$ | $t_0 = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$ | $t_0 < -t_{\alpha, v}$ |
| $H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$ | | $t_0 > t_{\alpha, v}$ |
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ | $v = n_1 + n_2 - 2$ para $\sigma_1^2 = \sigma_2^2$ $t_0 = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ | $ t_0 > t_{\alpha/2, v}$ |
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$ | | $t_0 < -t_{\alpha, v}$ |
| $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$ | o bien para $\sigma_1^2 \neq \sigma_2^2$ $v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$ $t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ | $t_0 > t_{\alpha, v}$ |

La varianza combinada se obtiene de la siguiente forma:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Tabla 2-3 Pruebas de varianza de distribuciones normales.

| HIPÓTESIS | ESTADÍSTICO DE PRUEBA | CRITERIO P/RECHAZO |
|---|--|---|
| $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \diamond \sigma_0^2$ | $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$ | $\chi_0^2 > \chi_{\alpha/2, n-1}^2$ o bien $\chi_0^2 > \chi_{1-\alpha/2, n-1}^2$ |
| $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$ | | $\chi_0^2 < \chi_{1-\alpha, n-1}^2$ |
| $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$ | | $\chi_0^2 > \chi_{\alpha, n-1}^2$ |
| $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \diamond \sigma_2^2$ | $F_0 = \frac{S_1^2}{S_2^2}$ | $F_0 > F_{\alpha/2, n_1-1, n_2-1}$ o bien $F_0 < F_{1-\alpha/2, n_1-1, n_2-1}$ |
| $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$ | $F_0 = \frac{S_2^2}{S_1^2}$ | $F_0 > F_{\alpha, n_2-1, n_1-1}$ |
| $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$ | $F_0 = \frac{S_1^2}{S_2^2}$ | $F_0 > F_{\alpha, n_1-1, n_2-1}$ |

Habr  notado estimado lector, que la prueba estad stica F intermedia tambi n se compara con el criterio de mayor, ya que el estad stico es m s sensible al menor cambio en la cola larga de la distribuci n de Fisher.

NOTA: El expositor, generar  ejemplos ilustrativos de cada una de las pruebas de hip tesis sintetizadas en las tablas 2-2 y 2-3.

3. Diseño de un solo factor

En la unidad 1, le presentamos una panorámica general del Diseño y Análisis de Experimentos (DAE), ahora nos corresponde hacer el procedimiento metodológico en el caso de un solo factor, y podremos crear la generalización para cuando atacemos problemas de calidad en los que estén involucrados más de un factor.

Es importante recordar, tres supuestos en los procedimientos de análisis de varianza para diseño de experimentos:

1. **CRITERIO DE HOMOGENIZACIÓN:** Hasta no probar lo contrario, el comportamiento de los datos es normal.
2. **CRITERIO DE HOMOELASTICIDAD:** Hasta no probar lo contrario, además de considerarse el comportamiento normal de los datos, consideraremos que la dispersión o variabilidad que puedan tener los mismos son iguales, es decir, aún desconociendo el valor de la varianza, la estamos suponiendo igual para los subconjuntos de datos involucrados.
3. **CRITERIO DE INDEPENDENCIA DE ERRORES:** Hasta no probar lo contrario, los errores aleatorios por variabilidad, se consideran independientes (por lo que su covarianza vale cero, se verá posteriormente).

Estos criterios implican en síntesis, la normalización de los datos del experimento, en la práctica, existe un efecto de compensación por la Ley de los grandes números y por la tendencia central demostrada en los experimentos de Shewhart, por lo que de aquí en adelante consideraremos que se cumplen en nuestras observaciones, muestreos y réplicas obtenidas.

Otra consideración a tomar en cuenta, es el aseguramiento por parte de nosotros que la secuencialización de toma de datos **ESTÁ LIBRE DE TENDENCIAS**, es decir, debemos de evitar que situaciones ajenas al factor y sus niveles de estudio del modelo estadístico se presente modificado por variables no controladas del experimento, si aún cuidando que esté libre de tendencias, el propio experimento la sugiere, entonces rediseñaremos el modelo y lograremos separar la causa que provoca dicha tendencia.

Un ejemplo de tendencia que reafirma el párrafo anterior es, digamos un equipo que tiene baja repetibilidad y se fuerza a seguir midiendo a las mismas condiciones la variable que hallamos definido como respuesta, éste tiende a desajustarse.

En un equipo que mide la dureza de materiales metálicos y usa una aguja con una cierta presión para que midamos como variable de respuesta la profundidad de la marca hecha en la probeta metálica, la aguja tiende a deformarse en el uso repetitivo sobre el mismo material.

Para reducir los efectos provocados por tendencias, debemos de secuencializar al azar las tomas de muestra y frecuentemente verificar el ajuste de nuestros equipos de medición, así como verificar que los datos están tomados a las condiciones físicas y humanas que realmente estamos fijando.

EJEMPLO DE SECUENCIALIZACIÓN: Se desea estudiar la resistencia a la tensión del cemento tipo Portland, para ello se tienen 3 técnicas de mezclado (1, 2 y 3), se ha determinado que el equipo de medición soporta correctamente 3 ensayos y debe verificarse y ajustarse para las siguientes mediciones, se diseña el experimento para tomar 5 datos de cada técnica, para evitar variaciones en la tensión registrada debidas al equipo de medición, se numeran los datos que se obtendrán de la siguiente forma:

Secuencia para obtener los datos de las técnicas de mezclado del cemento

| 1 | 2 | 3 |
|---|----|----|
| 1 | 6 | 11 |
| 2 | 7 | 12 |
| 3 | 8 | 13 |
| 4 | 9 | 14 |
| 5 | 10 | 15 |

Con un generador de números al azar, de 1 a 15, seleccionaremos el ensayo a llevar a cabo, teniendo en cuenta que cada 3 ensayos el equipo de medición necesita verificación y ajuste.

| No. Azar | Azar x 15 | Ensayo | Técnica | Lb/in ² |
|----------|-----------|--------|---------|--------------------|
| 0.341 | 5.115 | 6 | 2 | 3900 |
| 0.848 | 12.72 | 13 | 3 | 3800 |
| 0.509 | 7.635 | 8 | 2 * | 3800 |
| 0.680 | 10.2 | 11 | 3 | 4150 |
| 0.640 | 9.6 | 10 | 2 | 3950 |
| 0.795 | 11.925 | 12 | 3 * | 4000 |
| 0.307 | 4.605 | 5 | 1 | 4300 |
| 0.065 | 0.975 | 1 | 1 | 4100 |
| 0.904 | 13.56 | 14 | 3 * | 3900 |
| 0.113 | 1.695 | 2 | 1 | 4150 |
| 0.259 | 3.885 | 4 | 1 | 4000 |
| 0.977 | 14.655 | 15 | 3 * | 3950 |
| 0.181 | 2.715 | 3 | 1 | 4200 |
| 0.594 | 8.91 | 9 | 2 | 4000 |
| 0.461 | 6.915 | 7 | 2 | 4100 |

El asterisco (*) implica que después de ese ensayo deberá verificarse y ajustarse el equipo de medición, habiendo iniciado verificado y ajustado.

Ahora podemos proceder a ordenar conforme al esquema de ensayos:

TÉCNICAS DE MEZCLADO (Lb/in²)

| 1 | 2 | 3 |
|------|------|------|
| 4100 | 3900 | 4150 |
| 4150 | 4100 | 4000 |
| 4200 | 3800 | 3800 |
| 4000 | 4000 | 3900 |
| 4300 | 3950 | 3950 |

Utilizando la técnica de selección al azar, reducimos la posibilidad de incluir alguna tendencia de comportamiento en la toma de datos, ya sea por el experimentador o tendencia provocada por el equipo de medición.

En el ejercicio anterior, el factor en estudio es las TÉCNICAS DE MEZCLADO, en la cual como existen tres técnicas, tenemos tres niveles del factor, note estimado lector que específicamente están fijadas las tres técnicas, en estos casos les llamamos “modelo de efectos o niveles fijos”.

En oposición a la situación anterior, se da el caso muchas veces de tener un gran número de niveles (o efectos) y tener que escoger de ellos un cierto número manejable en el análisis de varianza, en estos casos se seleccionan al azar.

Supongamos que en el ejemplo de las técnicas de mezclado existieran, 50 técnicas y sólo debemos escoger 3 de ellas para hacer el estudio, en forma parecida a la secuencialización de la toma de datos, deberemos seleccionar primero a las técnicas.

- El primer número al azar que se obtuvo fue 0.381, éste lo multiplicamos por 50 y obtenemos 19.05, el criterio que hemos seguido es consistente a entero mayor por lo que en la primera columna colocaremos la técnica 20.
- El segundo No. es 0.750, por lo que nos da 37.5, entonces escogeremos la técnica 38.
- El tercero No. es 0.263, por lo que nos da 13.15, entonces escogeremos la técnica 14.

3.1 Modelo de efectos fijos

Los niveles del factor en estudio están fijos, en forma general es posible tomar diferente número de datos n_i , para el nivel (i) del factor, sea (a) el número de niveles por lo que para cada nivel se podrá tomar $j = 1, 2, 3, \dots, n_i$ datos tendremos un esquema siguiente:

| Factor A | Observaciones | | | |
|----------|---------------|----------|-------|-----------|
| 1 | X_{11} | X_{12} | | X_{1n1} |
| 2 | X_{21} | X_{22} | | X_{2n2} |
| : | : | : | | : |
| a | X_{a1} | X_{a2} | | X_{ana} |

En esta sección, desarrollaremos el procedimiento de análisis de varianza para el modelo de efectos fijos. Necesitaremos introducir alguna notación que nos permita identificar los totales y medias involucradas.

Sea T_i el total de las observaciones para el nivel i .

Sea \bar{X}_i la media en el nivel i .

Sus expresiones serían las siguientes:

$$T_i = \sum_{j=1}^{n_i} X_{ij} \quad i = 1, 2, \dots, a$$

$$\bar{X}_i = \frac{T_i}{n_i} \quad i = 1, 2, \dots, a$$

$$T_{..} = \sum_{i=1}^a \sum_{j=1}^{n_i} X_{ij}$$

$$\bar{X}_{..} = \frac{T_{..}}{N}$$

Partiendo de la diferencia de cada dato con la gran media de todos los datos y desglosando los niveles del factor llegamos:

$$\begin{array}{lcl} (X_{ij} - X_{..}) & = & (X_{i.} - X_{..}) + (X_{ij} - X_{i.}) \\ \text{Dif. de c/dato} & & \text{Dif. de c/nivel} \quad \text{Dif. no justificada} \\ \text{con la gran media} & & \text{con la gran media} \quad \text{genéricamente se le llama error} \end{array}$$

Obtenemos las ecuaciones de sumas de cuadrados:

$$SS_T = SS_A + SS_E$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2$$

$$SS_A = \sum_{i=1}^a n_i (\bar{X}_i - \bar{X}_{..})^2$$

$$SS_E = \sum_{i=1}^a \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

NOTA: El expositor desarrollará en clase los aspectos teóricos del desglose.

$$\bar{X}_{..} = \bar{X} \quad \text{y} \quad \bar{X}_{i.} = \bar{X}_i$$

El modelo estadístico para experimentos de solo un factor es:

$$X_{ij} = \mu + \alpha_i + e_{ij} \quad i = 1, 2, 3, \dots, a \quad j = 1, 2, 3, \dots, n_i \quad \text{Ec. 3-1}$$

En donde α_i representa el sesgo o desvío provocado por el nivel (i), la n_i representa el número de replicas que se encuentran en cada nivel, cuando el número de réplicas es diferente en cada nivel, generalmente decimos que el ANDEVA es desbalanceado, para problemas del área de Calidad se prefieren diseños balanceados, es decir el número de réplicas para cada nivel es igual a (n), hagamos ejemplos desbalanceados y balanceados.

DESBALANCEADOS:

EJEMPLO 3-1: Un metalúrgico probó la resistencia a la tensión de cinco aleaciones diferentes seleccionadas por el cliente, con varias muestras de cada aleación y los resultados fueron los siguientes:

| | | | | | | | T _{i.} | X _{i.} | T _{..} |
|-------------------|----------|------|------|------|------|------|-----------------|-----------------|-----------------|
| ALEACIONES | 1 | 12.4 | 19.8 | 15.2 | 14.8 | 18.5 | 80.7 | 16.14 | 309.8 |
| | 2 | 8.9 | 11.6 | 10.0 | 10.3 | | 40.8 | 10.20 | |
| | 3 | 10.5 | 13.8 | 12.1 | 11.9 | 12.6 | 60.9 | 12.18 | |
| | 4 | 12.8 | 14.2 | 15.9 | 14.1 | | 57.0 | 14.25 | X _{..} |
| | 5 | 16.4 | 15.9 | 17.8 | 20.3 | | 70.4 | 17.60 | 14.0818 |

- a) Calcular la resistencia media de cada aleación.
- b) Prepare un análisis de varianza de los datos, considerando que a mayor resistencia a la tensión es mejor, ¿existe diferencia estadística en las resistencias a la tensión de las aleaciones?, utilice nivel de significación de 5%.

A primera intención, ya habiendo calculado el inciso a), vemos en la columna de medias de cada aleación, que la 5 y 1 son las mayores, pero esto no implica que estadísticamente sean diferentes a las demás, hasta que probemos con un ANDEVA.

Existen N=22 datos totales, teniendo los valores 5,4,5,4 y 4 las (n) de cada nivel o efecto de cada aleación. El modelo está representado en la ecuación 3-1, al desglosar el modelo obtenemos:

$$SS_T = SS_A + SS_E \quad SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} (X_{i,j} - \bar{X})^2 = 210.7127$$

$$SS_A = \sum_{i=1}^a n_i (\bar{X}_i - \bar{X})^2 = 149.1627$$

la suma de cuadrados de error la podemos obtener por diferencia del total menos el factor.

$$SS_E = SS_T - SS_A = 61.55$$

Los grados de libertad son:

TOTALES = N - 1
FACTOR = a - 1
ERROR = N - a

Ahora procederemos a elaborar la tabla de ANDEVA.

| F. de V. | SS | GL | CM | F ₀ | F _{α, v₁, v₂} | DEC. | Valor Esp. |
|---------------------|----------|----|---------|----------------|--|------|--|
| Factor A en estudio | 149.1627 | 4 | 37.2907 | 10.2996 | 2.96 | SI | $\sigma^2 + \frac{\sum n_i \alpha_i^2}{a - 1}$ |
| Error | 61.55 | 17 | 3.6206 | | | | σ^2 |
| Total | 210.7127 | 21 | ----- | ----- | | | |

Al ser mayor el estadístico que la F de las tablas, la decisión (DEC.) fue afirmativa e implica que si existe suficiente razón estadística para concluir que las aleaciones provocan DIFERENTE RESISTENCIA A LA TENSIÓN.

Como es significativo el resultado, podemos calcular los valores alfa del modelo y reforzar nuestra decisión de que la ALEACIÓN 5 y en segundo lugar la aleación 1 son las más resistentes a la tensión.

$$\alpha_i = (X_{i.} - X_{..})$$

| | |
|------------------|-----------------|
| Valor de alfa 1: | 2.0582 |
| Valor de alfa 2: | - 3.8818 |
| Valor de alfa 3: | - 1.9018 |
| Valor de alfa 4: | 0.1682 |
| Valor de alfa 5: | 3.5182 |

En este ejemplo resulta evidente el resultado, sin embargo, no podíamos asegurarlo hasta tener la base metodológica del ANDEVA.

EJEMPLO 3-2:

Una organización de consumidores tiene interés en determinar si hay alguna diferencia en la vida promedio de cuatro marcas específicas de pilas del tipo "AA", habiéndose probado se obtuvo los siguientes resultados en horas.

| | | T _{i.} | X _{i.} | T _{..} | |
|--|---|-----------------|-----------------|-----------------|-----------------------|
| Marcas de Pilas "AA" Factor (A) en estudio | 1 | 15.4 | 16.1 | 15.2 | 16.3 |
| | 2 | 16.2 | 15.7 | 15.4 | |
| | 3 | 15.7 | 15.8 | 15.7 | 15.6 |
| | 4 | 15.9 | 15.5 | 15.5 | 16.1 |
| | | | | | X_{..} |

- ¿Existe diferencia significativa en las vidas promedios de las cuatro marcas específicas que se encuentran en estudio?, a mayor vida promedio es mejor.
- Hacer el ANDEVA y dar sus conclusiones pertinentes. Nivel de significación del 10%.

La tabla ANDEVA es la siguiente:

| F. de V. | SS | GL | CM | F ₀ | F _{α,ν₁,ν₂} | DEC. | Valor Esp. |
|---------------------|----|----|-------|----------------|--|------|--|
| Factor A en estudio | | | | | | | $\sigma^2 + \frac{\sum n_i \alpha_i^2}{a - 1}$ |
| Error | | | | | | | σ^2 |
| Total | | | ----- | ----- | | | |

BALANCEADO:

EJEMPLO 3-3:

Se utilizan dos procesos diferentes de templado, uno con agua salada y otro con aceite, en muestras de un tipo particular de aleación metálica. Los resultados se muestran a continuación. Considere que la dureza (o grado de temple) está distribuida normalmente.

FACTOR (A) EN ESTUDIO: Dureza obtenida.

NIVEL 1: Templado en agua salada.

NIVEL 2: Templado en aceite.

VARIABLE DE RESPUESTA: Fuerza al doblés en lb/in².

| Dureza | | | | | | | | | | | Ti. | Xi. | T.. |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 145 | 150 | 153 | 148 | 141 | 152 | 146 | 154 | 139 | 148 | | | |
| 2 | 152 | 150 | 147 | 155 | 140 | 146 | 158 | 152 | 151 | 143 | | | |

La dureza promedio es: X.. =

- Hacer el ANDEVA correspondiente. El nivel de significación es 5%.
- Dar sus conclusiones de ambos Templados.

| F. de V. | SS | GL | CM | F ₀ | F _{α,ν₁,ν₂} | DEC. | Valor Esp. |
|---------------------|----|----|-------|----------------|--|------|--|
| Factor A en estudio | | | | | | | $\sigma^2 + \frac{n \sum \alpha_i^2}{a - 1}$ |
| Error | | | | | | | σ^2 |
| Total | | | ----- | ----- | | | |

Cuando los niveles en estudio del factor son seleccionados al azar de un gran número de niveles posibles a escoger, este mismo procedimiento se utiliza pero las decisiones que se lleguen a tomar del ANDEVA, las consideraremos válidas para la generalidad de niveles.

La siguiente sección estudia este tipo de análisis.

3.2 Modelo de efectos al azar (aleatorizados)

El modelo estadístico (Ec. 3-1) es el mismo que el de la sección 3.1, y el procedimiento del ANDEVA también, pero si aplicamos el valor esperado al caso al AZAR la situación de haber seleccionado los niveles al azar provoca que el sesgo sea una varianza provocada por los niveles: σ_{α}^2 , representa la varianza por haber seleccionado los niveles al azar y como en cada nivel en estudio se toman (n_i , en el caso balanceado), el sesgo total del valor esperado del cuadrado medio del factor es $\sigma^2 + n\sigma_{\alpha}^2$.

Para el caso desbalanceado, el valor esperado es el mismo, pero se sustituye la n por n_0 , en donde el cálculo de n_0 está dado por la siguiente expresión:

$$n_0 = \frac{1}{a - 1} \left[\sum_{i=1}^a n_i - \frac{\sum_{i=1}^a n_i^2}{\sum_{i=1}^a n_i} \right]$$

3.3 Modelos por bloques

Al estar estudiando un factor, debemos cuidar que situaciones ajenas (no controladas) del experimento no estén presentes, ya que éstas se involucran en la suma de cuadrados del residuo (error), y por consecuencia las decisiones que tomemos serán inválidas, para evitar esta situación, desglosamos el modelo en bloques homogéneos (unidades experimentales), antes del desarrollo del experimento para posteriormente hacer la captura de datos.

El modelo de un factor por bloques es el siguiente: $X_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$

En donde α_i representa el Factor en estudio, la β_j representa el efecto del bloque y e_{ij} es la diferencia no justificada por el modelo (residuo o error aleatorio).

La ecuación de suma de cuadrados es la siguiente: $SS_T = SS_A + SS_B + SS_E$

EJEMPLO 3-4:

Una química desea probar el comportamiento de 4 agentes químicos, sobre la resistencia a la ruptura de un tipo particular de tela (cabeza de indio). Analiza las posibles variaciones que pueden estar presentes y encuentra que entre rollo de tela y otro rollo, puede existir variabilidad, si éste experimento se hiciera únicamente de un factor, ésta variabilidad no controlada se vería involucrada en la suma de cuadrados del error, aumentando su valor y la decisión tomada será errónea.

Para evitar la situación anterior, la química decide que en cada rollo seleccionado (tomará 5 son fijos) hará la prueba de cada uno de los 4 químicos (son fijos), así, cualquier variación provocada por los rollos no estará reflejada en el error. Realiza el experimento y considera un nivel de significación de 5%.

| FACTOR (A) | Rollos de tela | | | | | | |
|-----------------|----------------|-------|-------|-------|-------|------------------------|-------------------------|
| A. Químico | 1 | 2 | 3 | 4 | 5 | T _{i.} | X _{i.} |
| 1 | 73 | 68 | 74 | 71 | 67 | 353 | 70.60 |
| 2 | 73 | 67 | 75 | 72 | 70 | 357 | 71.40 |
| 3 | 75 | 68 | 78 | 73 | 68 | 362 | 72.40 |
| 4 | 73 | 71 | 75 | 75 | 69 | 363 | 72.60 |
| T _{.j} | 294 | 274 | 302 | 291 | 274 | T _{..} = 1435 | |
| X _{.j} | 73.50 | 68.50 | 75.50 | 72.75 | 68.50 | | X _{..} = 71.75 |

Las sumas de cuadrados se obtiene a partir del desglose de diferencias:

$$\begin{aligned}
 (X_{ij} - X_{..}) &= (X_{i.} - X_{..}) + (X_{.j} - X_{..}) + (X_{ij} - X_{i.} - X_{.j} + X_{..}) \\
 \text{Dif. de} & \quad \text{Dif. de} & \quad \text{Dif. de} & \quad \text{Dif. no justificada} \\
 \text{c/dato} & \quad \text{c/nivel} & \quad \text{c/bloque} & \quad \text{por el modelo} \\
 \text{con el total} & \quad \text{con el total} & \quad \text{con el total} & \quad
 \end{aligned}$$

Sus fórmulas correspondientes son:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n (X_{ij} - X_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n X_{ij}^2 - \frac{T_{..}^2}{an}$$

$$SS_A = \sum_{i=1}^a n(X_{i.} - X_{..})^2 = \frac{\sum_{i=1}^a T_i^2}{n} - \frac{T_{..}^2}{an}$$

$$SS_B = \sum_{j=1}^n a(X_{.j} - X_{..})^2 = \frac{\sum_{j=1}^n T_{.j}^2}{a} - \frac{T_{..}^2}{an}$$

$$SS_E = SS_T - SS_A - SS_B$$

Los valores esperados dependen si los bloques son fijos o bloques al azar y el factor fijo o al azar.

FACTOR Y BLOQUES FIJOS:

$$E(CM_A) = \sigma^2 + \frac{n \sum_{i=1}^a \alpha_i^2}{a - 1}$$

$$E(CM_B) = \sigma^2 + \frac{a \sum_{j=1}^n \beta_j^2}{n - 1}$$

$$E(CM_E) = \sigma^2$$

FACTOR Y BLOQUES AL AZAR:

$$E(CM_A) = \sigma^2 + n\sigma_\alpha^2$$

$$E(CM_B) = \sigma^2 + a\sigma_\beta^2$$

Continuando con el problema de la química (ejemplo 3-4), desarrollemos el ANDEVA correspondiente, recordando que la química estudia el factor (A), que son agentes químicos.

Los grados de libertad para cada fuente de variación serán los siguientes:

Total: $GL_T = an - 1 = N - 1$ ya que $N = an$
Factor A: $GL_A = a - 1$ ya que se tienen (a) niveles
Bloques: $GL_B = n - 1$ ya que se tienen (n) bloques
Error residual: $GL_E = (a - 1)(n - 1)$

| F. de V. | SS | GL | CM | F ₀ | F _{α, v₁, v₂} | Decisión |
|----------|--------|----|---------|----------------|--|----------|
| Factor A | 12.95 | 3 | 4.3167 | 2.3761 | 3.49 | NO |
| Bloques | 157.00 | 4 | 39.2500 | 21.61?* | 3.26* | ¿SI?* |
| Error | 21.80 | 12 | 1.8167 | | | |
| Total | 191.75 | 19 | ----- | ----- | | |

(*) La prueba de bloque sólo es indicativa, si resulta ser grande como en el ejemplo, nos guía a pensar que la variabilidad de los rollos de tela fue correcta que se restara al error (ver pags. 122, 123 y 126. "Diseño y Análisis de Experimentos", Montgomery).

Concluyendo, los agentes químicos no están siendo significativos para cambiar la resistencia a la ruptura de la tela, ya que no logramos rechazar.

3.4 Diseño Cuadrado Latino

En la sección 3.3 se analizó el modelo bloque completo, como un diseño para reducir el error de los residuos en un experimento, al sustraer la variabilidad debida a las unidades experimentales.

El cuadrado latino se utiliza para eliminar dos fuentes de variabilidad problemáticas, tienen un arreglo cuadrado y se utilizan las letras latinas (A,B,C,D,E, ... ETC.) para identificar las alternativas, efectos o tratamientos en estudio, equivale a desarrollar un análisis de bloque en dos direcciones para reducir la variabilidad del residuo o error.

El modelo estadístico para el cuadrado latino es el siguiente:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk}$$

En donde α_i identifica al bloque renglón con $i = 1, 2, \dots p$.

En donde β_j identifica al bloque columna con $j = 1, 2, \dots p$.

Y γ_k identifica los efectos o tratamientos con $k = 1, 2, \dots p$.

Los bloques renglón y columna son las fuentes de variabilidad "problemáticas" a tratar y el factor o tratamiento en estudio es la gamma. En los cuadrados latinos es necesario que haya "p" condiciones de cada una de las fuentes de variación.

Por lo que al crear un cuadrado latino generará $p \times p$ celdas, cada una de las p^2 , contendrá una de las p letras latinas que corresponde a un tratamiento del factor en estudio gamma y cada letra aparece una sola vez en cada renglón y columna correspondiente.

Como existen diferentes posibilidades de acomodar las p letras en arreglos de $p \times p$ entonces definiremos inicialmente **los cuadrados latinos estándar**.

Un cuadrado latino estándar es aquel que en su primer renglón el orden de la letras es alfabético y los renglones siguientes se escriben como el anterior desplazando un lugar hacia la izquierda.

Ejemplos de cuadrados latinos estándares:

| | 3x3 | 4x4 | 5x5 | 6x6 | 7x7 | p x p |
|-----------------------------------|------------|------------|---------------|------------------|-----------------------|--|
| | ABC | ABCD | ABCDE | ABCDEF | ABCDEFG | ABC...P |
| | BCA | BCDA | BCDEA | BCDEFA | BCDEFGA | BCD...A |
| | CAB | CDAB | CDEAB | CDEFAB | CDEFGAB | CDE...B |
| | | DABC | DEABC | DEFABC | DEFGABC | : |
| | | | EABCD | EFABCD | EFGABCD | : |
| | | | | FABCDE | FGABCDE | PAB... (P-1) |
| | | | | | GABCDEF | ----- |
| No. de Cuad. Estándares | 1 | 4 | 56 | 9408 | 16942080 | ----- |
| No. total de Cuad. Latinos | 12 | 576 | 161280 | 818851200 | 61479419904000 | $p!(p-1)!$ Cuadrados estándares. |

EJEMPLO 3-5: Suponga que un experimentador está estudiando el efecto de cinco fórmulas diferentes de la mezcla de dinamita sobre la fuerza explosiva observada. Cada fórmula se prepara usando un lote de materia prima, lo suficientemente grande para que sólo se logren hacer 5 mezclas. Las mezclas la preparan varios operadores, pudiendo existir una diferencia sustancial en la habilidad y experiencia entre ellos. Al parecer hay dos efectos extraños que se deben "separar" en el diseño que deseamos medir la fuerza explosiva de la dinamita, estos son: LOTES DE MATERIA PRIMA y OPERADORES. El diseño apropiado para este problema consiste en probar cada fórmula una sola vez, utilizando cada lote de materia prima y que cada fórmula sea preparada exactamente una vez por cada uno de los 5 operadores.

Sea A, B, C, D y E las 5 mezclas de dinamita en estudio.
 Sean los lotes de materia prima 1,2,3,4 y 5.
 Sean los operarios 1,2,3,4 y 5.

Se realiza la experimentación y se obtienen los siguientes resultados:

| Mat.Prim. | Operadores | | | | | T _{i..} | X _{i..} |
|------------------|------------|------|------|------|------|------------------------|------------------|
| | 1 | 2 | 3 | 4 | 5 | | |
| 1 | A=24 | B=20 | C=19 | D=24 | E=24 | 111 | |
| 2 | B=17 | C=24 | D=30 | E=27 | A=36 | 134 | |
| 3 | C=18 | D=38 | E=26 | A=27 | B=21 | 130 | |
| 4 | D=26 | E=31 | A=26 | B=23 | C=22 | 128 | |
| 5 | E=22 | A=30 | B=20 | C=29 | D=31 | 132 | |
| T _{.j.} | 107 | 143 | 121 | 130 | 134 | T _{... = 635} | |
| X _{.j.} | | | | | | X _{... =} | |

| T _{..k} | Fuerza explosiva de la Dinamita | | | | |
|------------------|---------------------------------|-----|-----|-----|-----|
| | A | B | C | D | E |
| X _{..k} | 143 | 101 | 112 | 149 | 130 |
| | | | | | |

Note estimado lector, que aunque existe tres contadores i, j, k y cada uno cuenta hasta p, el contador k está sujeto sólo a las permutaciones de las “fuentes de variabilidad problemáticas”, y sólo existen **pxp** celdas.

La suma de cuadrados se obtiene a partir del desglose de diferencias:

$$(X_{ijk} - X_{...}) = (X_{i..} - X_{...}) + (X_{.j.} - X_{...}) + (X_{..k} - X_{...}) + (X_{ijk} - X_{i..} - X_{.j.} - X_{..k} + 2X_{...})$$

Dif. de c/dato con el total Dif. Bloque Renglón Dif. Bloque Columna Dif. Efecto en estudio Dif. no justificada por el modelo

La suma de cuadrados es la siguiente:

$$SS_T = SS_{Renglón} + SS_{Columna} + SS_{Factor} + SS_E$$

Los grados de libertad son los siguientes:

| | |
|--------------------------|--|
| Bloque Renglón | p - 1 |
| Bloque Columna | p - 1 |
| Factor en estudio | p - 1 |
| Error | (p - 2)(p - 1) |
| Total | N - 1 en donde N = p² |

Las fórmulas obtenidas por el desglose de sumas de cuadrados son:

$$SS_T = \sum_i \sum_j \sum_k (X_{ijk} - X_{...})^2 = \sum_i \sum_j \sum_k X_{ijk}^2 - \frac{T_{...}^2}{N}$$

$$SS_{\text{Renglón}} = \sum_i p(X_{i..} - X_{...})^2 = \sum_{i=1}^p \frac{T_{i..}^2}{p} - \frac{T_{...}^2}{N}$$

$$SS_{\text{Columna}} = \sum_j p(X_{.j.} - X_{...})^2 = \sum_{j=1}^p \frac{T_{.j.}^2}{p} - \frac{T_{...}^2}{N}$$

$$SS_{\text{Factor}} = \sum_k p(X_{..k} - X_{...})^2 = \sum_{k=1}^p \frac{T_{..k}^2}{p} - \frac{T_{...}^2}{N}$$

$$SS_E = SS_T - SS_{\text{Renglón}} - SS_{\text{Columna}} - SS_{\text{Factor}}$$

Por lo que para el ejercicio de la dinamita tenemos los valores de las sumas de cuadrados, cuadrados medio, estadístico y decisión en la siguiente forma:

| F. de V. | SS | GL | CM | F ₀ | F _{α,v1,v2} | Decisión |
|----------------|--------|----|---------|----------------|----------------------|-----------------|
| Renglón | 68.00 | 4 | 17.00 | 1.59 | 3.26 | no al 5% |
| Columna | 150.00 | 4 | 37.50 | 3.51 | 3.26 | si al 5% |
| Factor | 330.00 | 4 | 82.50 | 7.73 | 3.26 | SI al 5% |
| Error | 128.00 | 12 | 10.67 | | | |
| Total | 676.00 | 24 | XXXXXXX | | | |

Observamos que el factor en estudio si fue significativo al 5% (si se hace la comparación al 1% también lo es:), por lo que concluimos que la fuerza explosiva de las formulaciones de dinamita, SI SON DIFERENTES ENTRE SI.

Como guía indicativa estimado lector, podemos notar que el efecto producido por los operarios, si sobrepasa el valor de las tablas correspondientes, esto nos confirma que un modelo de cuadrado latino fue adecuado para este experimento.

Para una mejor comparación en el estudio se determinan los valores estimados de los componentes gamma del modelo:

$$\bar{\gamma}_k = (X_{..k} - X_{...})$$

| Gamma 1 (A) | Gamma 2 (B) | Gamma 3 (C) | Gamma 4 (D) | Gamma 5 (E) |
|-------------|-------------|-------------|-------------|-------------|
| 3.20 | - 5.20 | - 3.00 | 4.40 | 0.60 |

Si la variable de respuesta es mejor mientras mayor sea en este ejemplo, la mezcla de dinamita (D) es la recomendable ya que tiene mayor fuerza explosiva.

En el caso en que accidentalmente “perdemos, o no pudimos obtener” un dato para tener balanceado el modelo por bloque (sección 3.3) o en el caso de cuadrado latino (sección 3.4), perdemos la ortogonalidad del factor en estudio con los bloques deberemos de proceder en la siguiente forma:

3.5 Estimación de valor faltante

En el caso de modelos por bloques (sección 3.3) existen dos formas de estimar el valor faltante, nosotros en este curso veremos la forma aproximada, que para usos prácticos en análisis de áreas de calidad es totalmente útil.

Si X_{ij} lo representamos como (x) , el dato faltante en el nivel (i) , bloque (j) , $T_{..}$ representa el gran total, $T_{i.}$ el total del nivel (i) , $T_{.j}$ el total del bloque (j) , estos totales son con un dato faltante, desarrollando la expresión de SS_E , derivando con respecto a (x) e igualando a cero obtenemos:

$$x = \frac{aT_{i.} + nT_{.j} - T_{..}}{(a - 1)(n - 1)}$$

Al obtener la estimación del valor faltante, se procede en la forma habitual de hacer el ANDEVA, sólo hay que cuidar que los grados de libertad totales cuando se estima algún valor faltante es un grado menos que si tuviéramos todos los datos completos.

En el caso de modelo cuadrado latino (sección 3.4), los totales están bajo las mismas condiciones obtenidos (sin el dato faltante), haciendo un procedimiento semejante obtenemos la estimación del dato faltante X_{ijk} también como (x) con la siguiente expresión:

$$x = \frac{p(T_{i..} + T_{.j.} + T_{..k}) - 2T_{...}}{(p - 2)(p - 1)}$$

3.6 Contrastes Ortogonales

En muchas situaciones en donde hacemos comparaciones múltiples, podemos llegar a tomar la decisión en el ANDEVA que existe diferencia significativa en las medias de los niveles en estudio, las siguientes preguntas que naturalmente debemos hacernos son:

- ¿Qué tan diferentes son las medias?
- ¿Qué medias difieren entre sí?
- ¿El promedio de un grupo de medias difiere con otro grupo de medias?

Estas preguntas podemos contestarlas creando “contrastes de comparación” entre los resultados obtenidos en el ANDEVA.

CONTRASTE: Implica tener una combinación lineal de los totales de los niveles, en el cual para el caso balanceado deberá estar sujeto que la suma de los coeficientes del contraste en los (a) niveles sumen cero.
En el caso desbalanceado, la combinación lineal de los totales de los niveles estarán sujetos a que la suma ponderada de los coeficientes del contraste sumen cero.

Si (C) es el contraste, y (c_i) es cada uno de los coeficientes del contraste en los (i) niveles que se tienen en estudio, entonces el contraste C estará dado por:

$$C = \sum_{i=1}^a c_i T_i. = n \sum_{i=1}^a c_i X_i. = \sum_{i=1}^a n_i c_i X_i.$$

En donde se presenta el caso general, el caso balanceado y el caso desbalanceado respectivamente; cada contraste está sujeto a la restricción, en los casos general y balanceado que la suma de los coeficientes del contraste sumen cero y en el caso desbalanceado la suma ponderada de los coeficientes del contraste sumen cero.

$$\text{Gral. y balanceado} \Rightarrow \sum_{i=1}^a c_i = 0$$

$$\text{Desbalanceado} \Rightarrow \sum_{i=1}^a n_i c_i = 0$$

La suma de cuadrados del contraste tiene **UN GRADO DE LIBERTAD** con la siguiente expresión tanto del balanceado como del desbalanceado.

$$SS_C = \frac{C^2}{n \sum_{i=1}^a c_i^2} = \frac{C^2}{\sum_{i=1}^a n_i c_i^2}$$

Para probar un contraste, este se debe comparar con el cuadrado medio del error, que en el caso de un factor, el estadístico tiene una distribución **F** con **1** y **N-a** grados de libertad.

CONTRASTES ORTOGONALES:

Quando se diseñan dos o más contrastes que cumplen que la suma de los productos de los coeficientes individuales sumen cero tenemos ortogonalidad entre los contrastes.

Para el caso desbalanceado, cada producto es ponderado con (n_i) datos del nivel (i).

Si C y D son contrastes ortogonales, entonces deberán cumplir con lo siguiente:

$$\text{Balanceado} \Rightarrow \sum_{i=1}^a c_i d_i = 0$$

$$\text{Desbalanceado} \Rightarrow \sum_{i=1}^a n_i c_i d_i = 0$$

Pueden obtenerse (**a-1**) contrastes ortogonales de un sistema de (**a**) niveles del factor, generalmente algo en la naturaleza del experimento nos debe sugerir las comparaciones que resultan de interés.

Como alternativas de contrastes veamos para $a=3$, $a=4$ y $a=5$.

| | Para a = 3: | | | Para a = 4: | | | | Para a=5: | | | | |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Contr. | T _{1.} | T _{2.} | T _{3.} | T _{1.} | T _{2.} | T _{3.} | T _{4.} | T _{1.} | T _{2.} | T _{3.} | T _{4.} | T _{5.} |
| C ₁ | -1 | 0 | +1 | +1 | 0 | 0 | -1 | -2 | -1 | 0 | +1 | +2 |
| C ₂ | +1 | -2 | +1 | 0 | +1 | -1 | 0 | +2 | -1 | -2 | -1 | +2 |
| C ₃ | | | | +1 | -1 | -1 | +1 | -1 | +2 | 0 | -2 | +1 |
| C ₄ | | | | | | | | +1 | -4 | +6 | -4 | +1 |

Fuente: Hicks pag. 324

Hicks pag. 33

Hicks pag. 324

EJEMPLO 3-6:

Se ha centrado el interés en el estudio de la conductividad medida para 4 revestimientos de tubos cinescopios para TV. Estos son específicos y de carácter cualitativo cada uno de ellos (I, II, III y IV). Se considera que 5 observaciones por revestimiento son adecuadas, por lo que se toman 20 cinescopios y se aplica un modelo de bloque completo al azar. El modelo estadístico es el planteado por la ecuación 3-1.

$$X_{ij} = \mu + \alpha_i + e_{ij}$$

Hay 4 niveles de tipo de revestimiento y 5 observaciones en cada uno, el objetivo general es seleccionar que revestimientos son los mejores, a menor valor de conductividad se es más aislante, aún seleccionando el mejor, por situaciones que tienen nuestros proveedores (inestabilidades en las entregas), las especificaciones de diseño en los cinescopios, debe ser menor o igual a 60, los datos son los siguientes:

| T. Rev. | | | | | | T _{i.} | X _{i.} |
|---------|----|----|----|----|----|-----------------|-----------------|
| I | 56 | 55 | 62 | 59 | 60 | 292.00 | 58.40 |
| II | 64 | 61 | 50 | 55 | 56 | 286.00 | 57.20 |
| III | 45 | 46 | 45 | 39 | 43 | 218.00 | 43.60 |
| IV | 42 | 39 | 45 | 43 | 41 | 210.00 | 42.00 |

T_{..} = 1006.00

X_{..} = 50.30

Al observar las medias de cada recubrimiento, los niveles 1 y 2 y por otro lado las medias 3 y 4 son cercanas, es de interés después de hacer el ANDEVA y si éste es rechazado (ser rechazado es ser significativamente diferentes), hacer contrastes para identificar los grupos de resultados convenientes para dirigir nuestros esfuerzos en el área de calidad para trabajar con nuestros proveedores y conjuntamente dar mejor producto a nuestros clientes finales.

El ANDEVA correspondiente a los revestimientos es el siguiente al 5% y al 1%:

| F. de V. | SS | GL | CM | F ₀ | F _{α,ν₁,ν₂} | DEC. | Valor Esp. |
|-------------------------|---------|----|--------|----------------|--|----------|--|
| Factor A Revestimiento. | 1135.00 | 3 | 378.33 | 29.79 | 3.24 al 5% 5.29 al 1% | SI SI | $\sigma^2 + \frac{n \sum \alpha_i^2}{a - 1}$ |
| Error | 203.20 | 16 | 12.70 | | | | σ^2 |
| Total | 1338.20 | 19 | ----- | ----- | | | |

Es altamente significativo que los 4 revestimientos se comportan sus conductividades en diferente forma en los cinescopios.

Determinaremos sus componentes del modelo (las alfas).

| Alfa 1 | Alfa 2 | Alfa 3 | Alfa 4 |
|--------|--------|--------|--------|
| + 8.10 | + 6.90 | - 6.70 | - 8.30 |

Observamos que existen dos grupos definidos por sus valores alfa, en las cuales se encuentran cercanos, un grupo es 1 y 2, el otro grupo es 3 y 4.

Nos interesa determinar si entre alfa 2 y alfa 3 estadísticamente difieren, por lo que desarrollaremos los siguientes contrastes.

| | T _{1.} | T _{2.} | T _{3.} | T _{4.} |
|----------------|-----------------|-----------------|-----------------|-----------------|
| C ₁ | 0 | +1 | -1 | 0 |
| C ₂ | +1 | 0 | 0 | -1 |
| C ₃ | -1 | +1 | +1 | -1 |

El contraste 1 es el que nos interesa mayormente y comparan lo siguiente:

| | | | |
|------------------|--|--|--|
| C ₁ : | H ₀ : α ₂ - α ₃ = 0 | H ₀ : α ₂ = α ₃ | H ₀ : μ ₂ = μ ₃ |
| C ₂ : | H ₀ : α ₁ - α ₄ = 0 | H ₀ : α ₁ = α ₄ | H ₀ : μ ₁ = μ ₄ |
| C ₃ : | H ₀ : (α ₂ + α ₃) - (α ₁ + α ₄) = 0 | H ₀ : μ ₂ + μ ₃ = μ ₁ + μ ₄ | |

A continuación procederemos a obtener el ANDEVA, desglosando los contrastes.

| F. de V. | SS | GL | CM | F ₀ | F _{α,ν₁,ν₂} | DEC. | Valor Esp. |
|-----------------------|---------|----|--------|----------------|--|----------|--|
| Factor Revestimiento. | 1135.00 | 3 | 378.33 | 29.79 | 3.24 al 5% 5.29 al 1% | SI SI | $\sigma^2 + \frac{n \sum \alpha_i^2}{a - 1}$ |
| C1 | 462.40 | 1 | 462.40 | 36.41 | 4.49 al 5% | SI | |
| C2 | 672.40 | 1 | 672.40 | 52.94 | 8.53 al 1% | SI | |
| C3 | 0.20 | 1 | 0.20 | 0.016 | | NO | |
| Error | 203.20 | 16 | 12.70 | | | | σ^2 |
| Total | 1338.20 | 19 | ----- | ----- | | | |

Los contrastes están sirviendo para desglosar aún más la suma de cuadrados del factor revestimiento (A). Si queremos contrastar los niveles que están visualmente quedando como grupo entonces el resultado sería el siguiente:

$$\begin{aligned} \mathbf{C1: (+1, -1, 0, 0)} & \quad C1 = (+1)(292) + (-1)(286) + (0)(218) + (0)(210) = \mathbf{6} \\ \mathbf{C2: (0, 0, +1, -1)} & \quad C2 = (0)(292) + (0)(286) + (+1)(218) + (-1)(210) = \mathbf{8} \\ \mathbf{C3: (+1, +1, -1, -1)} & \quad C3 = (+1)(292) + (+1)(286) + (-1)(218) + (-1)(210) = \mathbf{150} \end{aligned}$$

$$\mathbf{SS_1 = 3.60} \quad \mathbf{SS_2 = 6.40} \quad \mathbf{SS_3 = 1125} \quad \mathbf{SS_A = 1135} \quad \mathbf{SS_E = 203.20} \quad \mathbf{SS_T = 1338.2}$$

El grueso de las variaciones es provocado por la diferencia del grupo alto contra el bajo.

4. Diseños de más de un factor

En la sección anterior, hemos estudiado los esquemas más importantes que inciden en el estudio de solo un factor, ahora nos enfocaremos a los estudios simultáneos de dos o más factores.

Cuando tenemos más de un factor, estos generan situaciones que pueden presentarse, en forma general podemos tener:

1. **Experimentos Factoriales:** Son aquellos en los cuales tenemos presente todas las situaciones posibles de los niveles de dichos factores, el más simple es cuando existen dos factores, digamos **A** y **B**, en los cuales existen **a** y **b** niveles respectivamente, por lo que tenemos **ab** celdas totales de experimentación, si sólo se tiene una réplica por celda de experimentación tendremos mezclado el error experimental con la posible interacción que exista entre los factores, en este caso, deberemos asegurarnos antes del diseño y determinación del procedimiento de captura de datos que no exista interacción entre los niveles de **A** y los niveles de **B**.
2. **Experimentos anidados:** Son aquellos en los cuales un cierto factor, digamos **B**, se encuentra bajo las condiciones dadas de otro factor, digamos **A**, cada uno de los niveles de **B**, se encuentra condicionado (adentro) de cada uno de los niveles de **A**, de todas formas tendremos **ab** celdas de experimentación, pero difieren de la situación (1) puesto que los niveles de **B** no pueden separarse de los niveles de **A**. Nuevamente si en cada celda de experimentación sólo tenemos una réplica entonces estará mezclado los efectos producidos por el factor **B** con el error experimental.
3. **Experimentos Factoriales Anidados:** Son aquellos en los cuales se tiene presente las dos situaciones anteriores (1) y (2), tanto existe interacción como la situación de anidamiento, el caso más sencillo es cuando tenemos tres factores en estudio, digamos **A**, **B** y **C** con sus **a**, **b** y **c** niveles, en donde **B** se encuentra anidado dentro de **A** y **C** se encuentra cruzado factorialmente tanto con **A** como con **B**.

DEFINICIONES IMPORTANTES:

| | |
|----------------|---|
| n | Representa el número de datos (réplicas) en cada celda experimental. |
| celda | Representa al conjunto de datos que se encuentran a las mismas condiciones de los niveles en estudio. |
| efecto | Este término, se utilizará para expresar la modificación de valores en la variable de respuesta, tanto para los factores principales (A, B, C, ... etc.), así como sus interacciones (AB, AC, BC, ABC, ... etc.). |
| error | Se utilizará este término exclusivamente para denotar la variabilidad no justificada por el modelo, pero que proviene únicamente de la variabilidad experimental en la obtención de datos. |
| residuo | Se utilizará este término cuando no sea posible separar el efecto y el error, por lo que el resultado tiene mezcladas variabilidades. |

Es importante también referirlos a las “Reglas para sumas de cuadrados y Valores esperados de los cuadrados medios”, capítulo No. 8, pags. 229 a la 238; Diseño y Análisis de Experimentos, Montgomery. Ed. Iberoamérica.

Las reglas referidas en el párrafo anterior han sido analizadas por diversos autores, algunos son: Scheffé (1959), Bennett y Franklin (1954), Hicks (1973) y Searle (1971).

En estos apuntes estimado lector, se utilizará la letra griega Tau (τ) para representar interacciones cruzadas factoriales y se utilizará la letra latina (I) para representar interacciones del factor anidado con algún otro factor cruzado.

Así los modelos estadísticos factoriales y sus sumas de cuadrados son:

CASO A, B BIFACTORIAL:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \tau_{ij} + e_{ijk}$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

CASO A,B y C TRIFACTORIAL:

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \tau_{ij} + \tau_{ik} + \tau_{jk} + \tau_{ijk} + e_{ijkl}$$

$$SS_T = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_E$$

CASO B ANIDADO EN A:

$$X_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{(ij)k}$$

$$SS_T = SS_A + SS_{B(A)} + SS_E$$

CASO B ANIDADO EN A, FACTORIAL C:

$$X_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \tau_{ik} + I_{j(i)k} + e_{(ijk)l}$$

$$SS_T = SS_A + SS_{B(A)} + SS_C + SS_{AC} + SS_{B(A)C} + SS_E$$

4.1 Experimentos factoriales

En este tipo de plan, que consiste en tomar todos los factores principales y todas sus interacciones, el desglose de diferencias para el caso trifactorial es el siguiente:

$$\begin{aligned} (X_{ijkl} - X_{....}) &= (X_{i...} - X_{....}) + (X_{.j..} - X_{....}) + (X_{..k.} - X_{....}) + (X_{ij..} - X_{i...} - X_{.j..} + X_{....}) + \\ &+ (X_{i.k.} - X_{i...} - X_{..k.} + X_{....}) + (X_{.jk.} - X_{.j..} - X_{..k.} + X_{....}) + \\ &+ (X_{ijk.} - X_{ij..} - X_{i.k.} - X_{.jk.} + X_{i...} + X_{.j..} + X_{..k.} - X_{....}) + \\ &+ (X_{ijkl} - X_{ijk.}) \end{aligned}$$

Para determinar las sumas de cuadrados correspondientes, primeramente obtenemos los totales de cada nivel de factor y los totales que nos solicitan los contadores i, j y k para proceder a crear una tabla operacional en la cual obtendremos las sumas de cuadrados correspondientes.

EJEMPLO 4-1: En un cierto departamento de una compañía, trabajan dos ingenieros y tres operadores, y desarrollan un trabajo específico. El supervisor desea identificar si los ingenieros están trabajando igual, si no es el caso cual de ellos hace el trabajo en menor tiempo, también desea estudiar a los operadores en la misma forma y encontrar si entre pareja ingeniero-operador existe alguna interacción.

Para ello, diseña un experimento que por observación de ingeniero-operador hacen el mismo trabajo y mide el tiempo que se tardan en terminarlo, en condiciones habituales de ambiente de trabajo, cada pareja hace dos veces el trabajo y se obtiene la siguiente información:

FACTOR A: Ingenieros $i = 1, 2$ $a = 2$
FACTOR B: Operarios $j = 1, 2, 3$ $b = 3$
VARIABLE RESPUESTA: Tiempo en hrs.
No. DE RÉPLICAS: $n = 2$

| | Operador 1 | Operador 2 | Operador 3 | $T_{i..}$ | $X_{i..}$ |
|-------------|-------------|-------------|-------------|-------------------------------------|--------------------------------------|
| Ingeniero 1 | 2.59 | 2.48 | 2.50 | 15.19 | 2.53167 |
| | <u>2.61</u> | <u>2.49</u> | <u>2.52</u> | | |
| | 5.20 | 4.97 | 5.02 | | |
| Ingeniero 2 | 2.49 | 2.85 | 2.67 | 16.11 | 2.685 |
| | <u>2.58</u> | <u>2.72</u> | <u>2.80</u> | | |
| | 5.07 | 5.57 | 5.47 | | |
| $T_{.j.}$ | 10.27 | 10.54 | 10.49 | $T_{...} = 31.30$ | |
| $X_{.j.}$ | 2.5675 | 2.635 | 2.6225 | | $X_{...} = 2.6083$ |

Tabla de operaciones es:

| | $y = X_{ijk}$ | $y = T_{ij.}$ | $T_{i..}$ | $T_{.j.}$ | $T_{...}$ |
|----------------------|--------------------------|-----------------------------|--------------------------|--------------------------|-----------------|
| (1) Sumas y^2 | 81.8174 | 163.592 | 490.2682 | 326.6046 | 979.69 |
| (2) No. de Xs en c/y | 1 1 | n 2 | bn 6 | an 4 | abn 12 |
| (1)/(2) | 81.8174 | 81.796 | 81.71136 | 81.65115 | (3) 81.64083 |
| (1)/(2) - (3) | 0.176567 | 0.155167 | 0.070527 | 0.010317 | ----- |
| Sum de Cuad | SS_T | SS_{ST} | SS_A | SS_B | |

Si llamamos Suma de cuadrados subtotaes (ST), a la suma de A, B y AB, entonces por resta del total encontramos el error, por diferencia podemos encontrar la interacción AB.

$$SS_E = SS_T - SS_{ST} = 0.0214$$

$$SS_{AB} = SS_{ST} - SS_A - SS_B = 0.074323$$

Ahora procederemos al análisis de varianza:

| F. de V. | SS | GL | CM | F ₀ | F _{α, v₁, v₂} | DEC. | Valor Esp. |
|-------------------|----------|--------------|----------|----------------|--|----------|--|
| Factor A Ing. | 0.070527 | (a - 1) 1 | 0.070527 | 19.7739 | 5.99 al 5% 13.75 al 1 | SI SI | $\sigma^2 + \frac{bn \sum \alpha_i^2}{a - 1}$ |
| Factor B Oper. | 0.010317 | (b - 1) 2 | 0.005158 | 1.44617 | 5.14 al 5% 10.92 al 1 | NO NO | $\sigma^2 + \frac{an \sum \beta_j^2}{b - 1}$ |
| Inter. AB | 0.074323 | 2 | 0.037162 | 10.4191 | 5.14 al 5% 10.92 al 1 | SI NO | $\sigma^2 + \frac{n \sum \tau_{ij}^2}{(a - 1)(b - 1)}$ |
| Error | 0.0214 | 6 | 0.003567 | ----- | | | σ^2 |
| Total | 0.176567 | 11 | ----- | ----- | | | |

Es significativa la prueba para los Ingenieros aún al 1%, y también es significativa para la interacción al 5%, concluimos que el ingeniero 1 hace el mismo trabajo más rápidamente que el 2, los operadores no son significativas sus diferencias pero existe interacción y preferencias de los operadores con los ingenieros, obtengamos sus alfas y taus correspondientes.

$$\alpha_i = X_{i..} - X_{...}$$

$$\alpha_1 = -0.0766$$

$$\alpha_2 = +0.0766$$

$$\tau_{ij} = X_{ij.} - X_{i..} - X_{.j.} + X_{...}$$

$$\tau_{11} = +0.1091$$

$$\tau_{12} = -0.0733$$

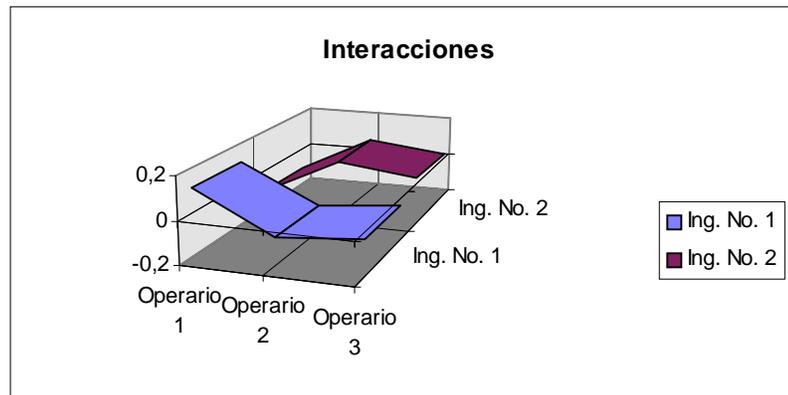
$$\tau_{13} = -0.0358$$

$$\tau_{21} = -0.1091$$

$$\tau_{22} = +0.0733$$

$$\tau_{23} = +0.0358$$

Observe estimado lector que los valores obtenidos suman cero tanto por renglón como por columna. El operador 1 trabaja mejor con el ingeniero 2, el operador 2 trabaja mejor con el ingeniero 1 y el operador 3 trabaja mejor con el ingeniero 1.



VALORES ESPERADOS: Para los casos en que tanto A como B sean seleccionados sus niveles al azar, ambos contribuyen en la varianza, si el modelo es mixto, es decir, A fijo y B al azar tenemos lo siguiente:

CASO 1: Ambos al azar:

$$\begin{aligned}
 E(CM_E) &= \sigma^2 \\
 E(CM_{AB}) &= \sigma^2 + n\sigma_{\alpha\beta}^2 \\
 E(CM_B) &= \sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_{\beta}^2 \\
 E(CM_A) &= \sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha}^2
 \end{aligned}$$

CASO 2: A fijo y B al azar:

$$\begin{aligned}
 E(CM_E) &= \sigma^2 \\
 E(CM_{AB}) &= \sigma^2 + n\sigma_{\alpha\beta}^2 \\
 E(CM_B) &= \sigma^2 + an\sigma_{\beta}^2 \\
 E(CM_A) &= \sigma^2 + n\sigma_{\alpha\beta}^2 + \frac{bn\sum\alpha_i^2}{a-1}
 \end{aligned}$$

El procedimiento es el mismo, pero tener cuidado en el planteamiento del estadístico a utilizar, debe ser conforme a los valores esperados.

4.2 Experimentos 2^k y técnica de Yates

En esta sección se presta atención especial a los diseños experimentales en los cuales el plan experimental tiene por objeto el estudio de (k) factores, los cuales cada uno de ellos tiene 2 niveles. Estos por lo común se conocen como experimentos 2^k factoriales frecuentemente los niveles se identifican como "Alto" y "Bajo" y se asocian con (+1) y (-1) respectivamente, también es muy frecuente que se manejen sólo por el signo (+) y (-).

Este plan es particularmente útil en las fases tempranas del trabajo experimental, ya que permite el estudio de muchos factores simultáneos, obviamente la capacidad de estudiar (k) factores con un método sencillo nos permite posteriormente profundizar nuestra atención a los factores de mayor incidencia en el trabajo experimental.

Utilizaremos una notación simplificada que representa el estado completo de la celda experimental:

| | |
|------------------------|---|
| n | Representa el número de réplicas en cada celda. |
| (1) | Representa que todos los factores se encuentran en su nivel bajo. |
| a,b,c, ... etc. | La presencia de una letra minúscula representa que ese factor se encuentra en su nivel alto. Existe un orden estándar de acomodo para el cálculo de los contrastes correspondientes. |

En un principio consideraremos el plan factorial (2^2) en el cual se tienen (n) observaciones experimentales o réplicas en cada celda, utilizando la notación de Yates interpretamos los símbolos **(1)**, **a**, **b**, **ab** como los totales por cada celda, podemos definir los contrastes para cada factor e interacción:

Contraste para A: $C_A = ab + a - b - (1)$ **Coef. = +1, +1, -1, -1**
Contraste para B: $C_B = ab - a + b - (1)$ **Coef. = +1, -1, +1, -1**
Contraste para AB: $C_{AB} = ab - a - b + (1)$ **Coef. = +1, -1, -1, +1**

| | B "bajo" (-1) | B "alto" (+1) | Media |
|---------------|----------------------|---------------------|-------------------------------|
| A "bajo" (-1) | (1) | b | $\frac{b + (1)}{2n}$ |
| A "alto" (+1) | a | ab | $\frac{ab + a}{2n}$ |
| Media | $\frac{a + (1)}{2n}$ | $\frac{ab + b}{2n}$ | $\frac{ab + a + b + (1)}{4n}$ |

Para obtener los diferentes contrastes cuando el plan sea a más de dos factores, es decir, sea (k) los factores que se estén manejando, Yates (1937) propone un orden sencillo, adecuado y operativo para su obtención.

Este consiste en inicialmente colocar la celda en que todos los factores estén en sus niveles bajos y se identifica con el símbolo **(1)**. Posteriormente se coloca el factor A en alto con los demás bajos (**a**), seguido por B alto (**b**) y (**ab**), veamos las siguientes secuencias estándares:

Factores A y B: **(1), a, b, ab**
Factores A, B y C: **(1), a, b, ab, c, ac, bc, abc**
Factores A, B, C y D: **(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd**

Cada factor adicional que se ordene hasta llegar a (k)ésimo factor, continúa la secuencia por multiplicación sucesiva con los anteriores, a este ordenamiento de le llama secuencia estándar de Yates.

Retomando el plan 2^2 y colocando las celdas experimentales en la secuencia estándar, la técnica de Yates plantea la ubicación de (k) columnas para la obtención de los contrastes:

En el caso de dos factores, podemos ayudarnos colocando tabla de signos y la obtención de los contrastes:

| A | B | Celda | Columna 1 | Columna 2 (Total y Contrastes) |
|----|----|------------|-----------|--------------------------------|
| -- | -- | (1) | $a + (1)$ | $ab + a + b + (1)$ |
| + | -- | a | $ab + b$ | $ab + a - b - (1)$ |
| -- | + | b | $a - (1)$ | $ab - a + b - (1)$ |
| + | + | ab | $ab - b$ | $ab - a - b + (1)$ |

TÉCNICA DE YATES PARA CÁLCULO DE CONTRASTES:

En verdad resulta laborioso elaborar la tabla de signos y las expresiones para los contrastes en experimentos grandes, por lo tanto Yates elaboró una técnica que consta de los siguientes pasos:

1. Se colocan las celdas y los totales por celda en columnas en ORDEN ESTÁNDAR.
2. Se identifica la mitad de los renglones y se generan (**k**) columnas operativas de la siguiente manera:
3. Se obtiene la mitad superior de la primera columna sumando los primeros dos resultados, enseguida los dos siguientes, así sucesivamente hasta agotar los pares (habremos llenado la mitad superior). La mitad inferior se obtiene restando por pares el primero del segundo (segundo - primero).
4. Se repite el procedimiento del paso 3 hasta obtener la (**k**)ésima columna para un experimento factorial 2^k .
5. En la columna (**k**) habremos obtenido, en el primer renglón el gran total y en cada uno de los siguientes, los contrastes correspondientes, el efecto o diferencia promedio del nivel ("alto" - "bajo") lo obtenemos dividiendo el contraste entre $2^{(k-1)}n$ y la suma de cuadrados correspondiente la obtenemos con el contraste al cuadrado entre $2^k n$.

En el de dos factores los efectos son los siguientes:

Efecto de A: $(X_{2..} - X_{1..}) = (ab + a - b - (1))/(2n)$

Efecto de B: $(X_{.2.} - X_{.1.}) = (ab - a + b - (1))/(2n)$

Para el de tres factores tendremos:

| A | B | C | Celda | Columna 1 | Columna 2 | Columna 3 (Total y Contrastes) |
|----|----|----|-------|-----------|-------------|-----------------------------------|
| -- | -- | -- | (1) | a + (1) | ab+a+b+(1) | abc+ab+ac+bc+a+b+c+(1) |
| + | -- | -- | a | ab + b | abc+ac+bc+c | abc+ab+ac-bc+a-b-c-(1) |
| -- | + | -- | b | ac + c | ab+a-b-(1) | abc+ab-ac+bc-a+b-c-(1) |
| + | + | -- | ab | abc + bc | abc+ac-bc-c | abc+ab-ac-bc-a-b+c+(1) |
| -- | -- | + | c | a - (1) | ab-a+b-(1) | abc-ab+ac+bc-a-b+c-(1) |
| + | -- | + | ac | ab - b | abc-ac+bc-c | abc-ab+ac-bc-a+b+c+(1) |
| -- | + | + | bc | ac - c | ab-a-b+(1) | abc-ab-ac+bc+a-b-c+(1) |
| + | + | + | abc | abc - bc | abc-ac-bc+c | abc-ab-ac-bc+a+b+c-(1) |

EJEMPLO 4-2:

En un experimento metalúrgico se desea probar el efecto de 4 factores y sus interacciones sobre la concentración (porcentaje en peso) de un compuesto particular de fósforo en material vaciado. Las variables son **(A) porcentaje de fósforo en la refinación, (B) porcentaje de material reprocesado vuelto a fundir, (C) tiempo de fundición,** y el factor **(D) tiempo de contención.** Los cuatro factores se varían en un experimento factorial 2^4 con dos fundiciones en cada condición. Se llevan a cabo las 32 fundiciones en orden al azar y se obtienen los siguientes datos.

| Celda | Dato 1 | Dato 2 | Total | Col. 1 | Col. 2 | Col. 3 | Col. 4 | Efecto | SS |
|-------|--------|--------|-------|--------|--------|--------|--------|---------|---------------|
| (1) | 30.3 | 28.6 | 58.9 | 118.8 | 222.0 | 420.6 | 865.0 | XXXXX | XXXXX |
| a | 28.5 | 31.4 | 59.9 | 103.2 | 198.6 | 444.4 | -19.2 | -1.2 | 11.520 |
| b | 24.5 | 25.6 | 50.1 | 98.9 | 228.3 | 1.0 | -19.6 | -1.225 | 12.005 |
| ab | 25.9 | 27.2 | 53.1 | 99.7 | 216.1 | -20.2 | 15.8 | 0.9875 | 7.801 |
| c | 24.8 | 23.4 | 48.2 | 116.0 | 4.0 | -14.8 | -35.6 | -2.225 | 39.605 |
| ac | 26.9 | 23.8 | 50.7 | 112.3 | -3.0 | -4.8 | 9.8 | 0.6125 | 3.001 |
| bc | 24.8 | 27.8 | 52.6 | 108.6 | -18.5 | -6.0 | 19.0 | 1.1875 | 11.281 |
| abc | 22.2 | 24.9 | 47.1 | 107.5 | -1.7 | 21.8 | -8.8 | -0.55 | 2.420 |
| d | 31.7 | 33.5 | 65.2 | 1 | -15.6 | -23.4 | 23.8 | 1.4875 | 17.701 |
| ad | 24.6 | 26.2 | 50.8 | 3 | 0.8 | -12.2 | -21.2 | -1.325 | 14.045 |
| bd | 27.6 | 30.6 | 58.2 | 2.5 | -3.7 | -7.0 | 10.0 | 0.625 | 3.125 |
| abd | 26.3 | 27.8 | 54.1 | -5.5 | -1.1 | 16.8 | 27.8 | 1.7375 | 24.151 |
| cd | 29.9 | 27.7 | 57.6 | -14.4 | 2.0 | 16.4 | 11.2 | 0.7 | 3.920 |
| acd | 26.8 | 24.2 | 51.0 | -4.1 | -8.0 | 2.6 | 23.8 | 1.4875 | 17.701 |
| bcd | 26.4 | 24.9 | 51.3 | -6.6 | 10.3 | -10.0 | -13.8 | -0.8625 | 5.951 |
| abcd | 26.9 | 29.3 | 56.2 | 4.9 | 11.5 | 1.2 | 11.2 | 0.7 | 3.920 |
| sumas | 428.1 | 436.9 | 865.0 | XXXXX | XXXXX | XXXXX | XXXXX | XXXXX | XXXXX |

La suma de cuadrados totales obtiene **217.509** y por diferencia se obtiene para el error lo siguiente: **39.362** por lo que podemos proceder al ANDEVA.

| F. de V. | SS | GL | CM | F ₀ | F _{α, v₁, v₂} | DEC. |
|----------|---------|----|--------|----------------|--|------|
| A | 11.520 | 1 | 11.520 | 4.683 | 4.49 al 5% | SI |
| B | 12.005 | 1 | 12.005 | 4.880 | | SI |
| C | 39.605 | 1 | 39.605 | 16.099 | | SI |
| D | 17.701 | 1 | 17.701 | 7.195 | | SI |
| AB | 7.801 | 1 | 7.801 | 3.171 | | NO |
| AC | 3.001 | 1 | 3.001 | 1.220 | | NO |
| AD | 14.045 | 1 | 14.045 | 5.709 | | SI |
| BC | 11.281 | 1 | 11.281 | 4.586 | | SI |
| BD | 3.125 | 1 | 3.125 | 1.270 | | NO |
| CD | 3.920 | 1 | 3.920 | 1.593 | | NO |
| ABC | 2.420 | 1 | 2.420 | 0.984 | | NO |
| ABD | 24.151 | 1 | 24.151 | 9.817 | | SI |
| ACD | 17.701 | 1 | 17.701 | 7.195 | | SI |
| BCD | 5.951 | 1 | 5.951 | 2.419 | | NO |
| ABCD | 3.920 | 1 | 3.920 | 1.593 | | NO |
| Error | 39.362 | 16 | 2.460 | XXXXX | | |
| Total | 217.509 | 31 | XXXXX | XXXXX | | |

Los factores A, B, C y D son significativos al 5%, las interacciones dobles AD y BC son significativas, las interacciones triples ABD y ACD son significativas, las demás no lo fueron.

CONCLUSIONES: Al ser cada factor individualmente significativo representa que los cuatro factores son críticos para el porcentaje en peso del compuesto de fósforo en el material vaciado.

CONCLUSIONES: (Continuación).

Las interacciones dobles: **AD y BC** resultaron significativas, las **AB, AC, BD y CD** no lo son.

Las interacciones triples **ABD y ACD** resultaron significativas, las demás interacciones triples no lo son. La interacción **ABCD** no resultó significativa.

Las anteriores descripciones conducen estimado lector que observando los efectos promedios de ("Alto - Bajo"), es decir, teniendo el proceso Metalúrgico de fundición en estado bajo en todos los factores (**1**) y trasladarlo a que todos sus factores estén en alto, **abcd** se produce un equilibrio interno de factores (homeóstasis), por esta razón la interacción **abcd** no fue significativa, su contraste correspondiente es:

$$C_{abcd} = abcd-abc-abd-acd-bcd+ab+ac+ad+bc+bd+cd-a-b-c-d+(1) = 11.2$$

Los efectos creados en las 32 fundiciones en estudio genera los siguientes cambios promedios:

| Efectos | Significativos | ("Alto-Bajo") | Efectos | No significativos |
|---------|----------------|---------------|---------|-------------------|
| A | - 1.2000 | | AB | 0.9875 |
| B | - 1.2250 | | AC | 0.6125 |
| C | - 2.2250 | | BD | 0.6250 |
| D | 1.4875 | | CD | 0.7000 |
| AD | - 1.3250 | | ABC | - 0.5500 |
| BC | 1.1875 | | BCD | - 0.8625 |
| ABD | 1.7375 | | ABCD | 0.7000 |
| ACD | 1.4875 | | | |

Habrà notado estimado lector, tanto en el ANDEVA como en los efectos promedios que nos señalan donde debemos abocar nuestra atención en los efectos, la suma de variaciones significativas son **148.009**, de un total de **217.509** y es con el siguiente orden:

| Lugar | Efecto | F ₀ | Valor del Efecto | %Acum.Sig | %Acum.Total |
|-------|--------|----------------|------------------|-----------|-------------|
| 1 | C | 16.099 | - 2.2250 | 26.76% | 18.21% |
| 2 | ABD | 9.817 | 1.7375 | 43.08% | 29.31% |
| 3 | D | 7.195 | 1.4875 | 55.04% | 37.45% |
| 3 | ACD | 7.195 | 1.4875 | 66.99% | 45.59% |
| 4 | AD | 5.709 | - 1.3250 | 76.48% | 52.05% |
| 5 | B | 4.880 | - 1.2250 | 84.59% | 57.56% |
| 6 | A | 4.683 | - 1.2000 | 92.38% | 62.86% |
| 7 | BC | 4.586 | 1.1875 | 100.00% | 68.05% |

El 53.33% de las 15 fuentes de variación fueron significativas, el tiempo de fundición **C**, la interacción **ABD** y **D** representan el 55.04% de las variaciones significativas de este problema, y representan el 20% de las fuentes de variación significativas.

4.3 Experimentos anidados

En muchos experimentos que contienen más de un factor, los niveles de algunos factores pueden estar condicionados a la presencia necesaria de un cierto nivel de otro factor, entonces podemos notar que existe jerarquización de un factor sobre otro factor, en estos casos el modelo a utilizar debe ser anidado.

En el modelo simple de **dos etapas** tenemos dos factores **A** y **B** en los cuales existen **(a)** niveles de **A** y **(b)** niveles de **B**, pero los niveles del factor **B** se encuentran condicionados a cada uno de los niveles de **A**, el nivel 1 de **B** dado el nivel 1 de **A**, es en este modelo claramente diferente al nivel 1 de **B** dado el nivel 2 de **A**, aunque pueden ser similares en algunos otros atributos.

El modelo y la suma de cuadrados es la siguiente:

$$X_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{(ij)k}$$

$$SS_T = SS_A + SS_{B(A)} + SS_E$$

EJEMPLO 4-3:

Como ejemplo ilustrativo de estas situaciones, considere que en una empresa se tienen dos áreas que se hace el mismo trabajo, digamos en diferentes localidades, en cada área se tiene una máquina y tres operarios que la trabajan, nuestro interés radica en estudiar el factor **(A) MÁQUINAS** y el factor **(B) OPERARIOS**, debido a que las áreas se encuentran en diferente localidad no es posible que los 6 operarios existentes trabajen en las dos máquinas, por lo que no puede ser el modelo factorial, deberemos jerarquizar en cada máquina a los tres operarios que trabajan en ella. Las áreas producen las piezas y son enviadas a nuestra central, un cierto atributo de calidad relacionado con el comportamiento de las máquinas y la habilidad de los operarios es medido en escala de 0 a 100, considerando especificación mínima de 65.

Se toma la decisión de que cada operario genere tres piezas para ser evaluadas con el criterio de calidad establecido (escala) y obtenemos los siguientes resultados:

| Factor (A) | Máquina 1 | | | Máquina 2 | | |
|---------------------|-----------|---------|---------|-----------|---------|---------|
| Factor (B) | Oper. 1 | Oper. 2 | Oper. 3 | Oper. 1 | Oper. 2 | Oper. 3 |
| | 69 | 68 | 71 | 69 | 72 | 72 |
| | 70 | 68 | 71 | 69 | 71 | 71 |
| | 69 | 69 | 70 | 69 | 72 | 71 |
| $T_{j(i) \cdot}$ | 208 | 205 | 212 | 207 | 215 | 214 |
| $T_{i \cdot \cdot}$ | 625 | | | 636 | | |
| T_{\dots} | 1261 | | | | | |

Las diferencias y las sumas de cuadrados tienen las siguientes expresiones:

$$(X_{ijk} - X_{...}) = (X_{i..} - X_{...}) + (X_{j(i).} - X_{i..}) + (X_{ijk} - X_{j(i).})$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - \frac{T_{...}^2}{abn} \quad \text{G.L.} = abn - 1$$

$$SS_A = \sum_{i=1}^a \frac{T_{i..}^2}{bn} - \frac{T_{...}^2}{abn} \quad \text{G.L.} = a - 1$$

$$SS_{B(A)} = \sum_{i=1}^a \sum_{j=1}^b \frac{T_{j(i).}^2}{n} - \sum_{i=1}^a \frac{T_{i..}^2}{bn} \quad \text{G.L.} = a(b - 1)$$

$$SS_E = SS_T - SS_A - SS_{B(A)} \quad \text{G.L.} = ab(n - 1)$$

Tabla de operaciones es:

| | $y = X_{ijk}$ | $y = T_{j(i).}$ | $T_{i..}$ | $T_{...}$ |
|----------------------|---------------|-----------------|-----------------|-----------------|
| (1) Sumas y^2 | 88371 | 265103 | 795121 | 1590121 |
| (2) No. de Xs en c/y | 1 1 | n 3 | bn 9 | abn 18 |
| (1)/(2) | 88371 | 88367.67 | (4) 88346.78 | (3) 88340.06 |
| (1)/(2) - (3o4) | 30.94 | 20.89 | 6.72 | ----- |
| Sum de Cuad | SS_T | $SS_{B(A)}$ | SS_A | |

Ahora procederemos al ANDEVA:

| F. de Var. | SS | GL | CM | F_0 | F_{α, v_1, v_2} | DECISIÓN |
|------------|-------|----|--------|---------|------------------------|----------|
| A Maq's | 6.72 | 1 | 6.7200 | 24.2162 | 4.75 al 5% | SI |
| B(A) Oper. | 20.89 | 4 | 5.2225 | 18.8198 | 3.26 al 5% | SI |
| Error | 3.33 | 12 | 0.2775 | XXXXXX | XXXXXX | XXXXXX |
| Total | 30.94 | 17 | XXXXXX | XXXXXX | XXXXXX | XXXXXX |

Note estimado lector que son significativos tanto las máquinas como los operadores, y la comparación de los operadores está bajo la condición de la máquina en que operan. Calcularemos las alfas y betas correspondientes:

$$\begin{array}{lll} \alpha_1 = -0.611 & \beta_{1(1)} = -0.111 & \beta_{1(2)} = -1.666 \\ \alpha_2 = 0.611 & \beta_{2(1)} = -1.111 & \beta_{2(2)} = 1.000 \\ & \beta_{3(1)} = 1.222 & \beta_{3(2)} = 0.666 \end{array}$$

4.4 Experimentos Factorial-Anidado

En esta sección manejaremos experimentos en los cuales se presentan las situaciones analizadas individualmente en las secciones anteriores (4.1, 4.2 y 4.3), el caso más sencillo es cuando tenemos 3 factores en estudio digamos **A**, **B** y **C** en los cuales B se encuentra anidado en A y C se encuentra cruzado tanto a A como a B. El modelo, sus sumas de cuadrados provenientes de la diferencia y las fórmulas correspondientes se presentan a continuación.

$$X_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \tau_{ik} + I_{j(i)k} + e_{(ijk)l}$$

$$SS_T = SS_A + SS_{B(A)} + SS_C + SS_{AC} + SS_{B(A)C} + SS_E$$

$$(X_{ijkl} - X_{....}) = (X_{i...} - X_{....}) + (X_{j(i)...} - X_{i...}) + (X_{...k} - X_{....}) + \\ + (X_{i \cdot k} - X_{i...} - X_{...k} + X_{....}) + (X_{j(i)k} - X_{j(i)...} - X_{i \cdot k} + X_{i...}) + \\ + (X_{ijkl} - X_{j(i)k})$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n X_{ijkl}^2 - \frac{T_{....}^2}{abcn} \quad \text{G.L.} = abcn - 1$$

$$SS_A = \sum_{i=1}^a \frac{T_{i...}^2}{bcn} - \frac{T_{....}^2}{abcn} \quad \text{G.L.} = a - 1$$

$$SS_{B(A)} = \sum_{i=1}^a \sum_{j=1}^b \frac{T_{j(i)...}^2}{cn} - \sum_{i=1}^a \frac{T_{i...}^2}{bcn} \quad \text{G.L.} = a(b - 1)$$

$$SS_C = \sum_{k=1}^c \frac{T_{...k}^2}{abn} - \frac{T_{....}^2}{abcn} \quad \text{G.L.} = c - 1$$

$$SS_{AC} = \sum_{i=1}^a \sum_{k=1}^c \frac{T_{i \cdot k}^2}{bn} - \sum_{i=1}^a \frac{T_{i...}^2}{bcn} - \sum_{k=1}^c \frac{T_{...k}^2}{abn} + \frac{T_{....}^2}{abcn} \quad \text{G.L.} = (a-1)(c-1)$$

$$SS_{B(A)C} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \frac{T_{j(i)k}^2}{n} - \sum_{i=1}^a \sum_{j=1}^b \frac{T_{j(i)...}^2}{cn} - \sum_{i=1}^a \sum_{k=1}^c \frac{T_{i \cdot k}^2}{bn} + \sum_{i=1}^a \frac{T_{i...}^2}{bcn}$$

Los grados de libertad para la interacción B(A)C es: **G.L. = a(b-1)(c-1)**

EJEMPLO 4-4:

Un especialista en el área de calidad, ha desarrollado un nuevo procedimiento para la recarga de máquina en el proceso de elaboración de envases de plástico, existen 3 diferentes zonas geográficas en donde se lleva a cabo el proceso y en cada zona se tienen 3 operarios que la manejan, no es posible hacer traslado de operarios entre las zonas y se desean probar tanto el procedimiento propuesto como el actual, en las tres zonas, el criterio de medición es el número de ciclos promedio por cada hora que logran obtener los operadores. Se realiza la experimentación tomando dos réplicas seleccionando al azar los días de medición y los resultados obtenidos en la semana pasada son:

| Zona> | 1 | | | 2 | | | 3 | | | |
|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|------------------|
| Oper> | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | T _{..k} |
| Proc. (1) | 20.2 24.1 | 26.2 26.9 | 23.8 24.9 | 22.0 23.5 | 22.6 24.6 | 22.9 25.0 | 23.1 22.9 | 22.9 23.7 | 21.8 23.5 | |
| T _{j(i)k} | 44.3 | 53.1 | 48.7 | 45.5 | 47.2 | 47.9 | 46.0 | 46.6 | 45.3 | |
| T _{i.k} | 146.1 | | | 140.6 | | | 137.9 | | | 424.6 |
| Proc. (2) | 20.0 19.8 | 21.4 21.7 | 20.4 20.5 | 19.9 20.1 | 20.7 20.9 | 18.7 18.9 | 17.9 18.0 | 18.9 19.0 | 17.5 17.9 | |
| T _{j(i)k} | 39.8 | 43.1 | 40.9 | 40.0 | 41.6 | 37.6 | 35.9 | 37.9 | 35.4 | |
| T _{i.k} | 123.8 | | | 119.2 | | | 109.2 | | | 352.2 |
| T _{j(i)..} | 84.1 | 96.2 | 89.6 | 85.5 | 88.8 | 85.5 | 81.9 | 84.5 | 80.7 | |
| T _{i...} | 269.9 | | | 259.8 | | | 247.1 | | | |
| T _{....} | 776.8 | | | | | | | | | |

La tabla de operaciones es la siguiente:

| | $y = X_{ijkl}$ | $y = T_{j(i)k}$ | $y = T_{i.k}$ | $y = T_{j(i)..}$ | T _{..k} | T _{i...} | T _{....} |
|-----------------------------|-----------------------|-------------------------|-------------------------|--------------------------|-----------------------|-----------------------|-------------------|
| (1) Sumas de y ² | 16977.34 | 33923.10 | 101589.70 | 67221.70 | 304330.00 | 201400.46 | 603418.24 |
| (2) No. de X en c/y | 1 1 | n 2 | bn 6 | cn 4 | abn 18 | bcn 12 | abcn 36 |
| (1)/(2) | 16977.3400 | 16961.5500 | 16931.6167 | 16805.4250 | 16907.2222 | 16783.3717 | 16761.6178 |
| 1/2- (3o4) | 215.7222 | 199.9322 | 169.9989 | 22.0533 | 145.6044 | 21.7539 | XXXXXX |
| | SS_T | SS_{ST1} | SS_{ST2} | SS_{B(A)} | SS_C | SS_A | |

En donde:

$$SS_T = SS_{ST1} + SS_E$$

$$SS_{ST1} = SS_A + SS_{B(A)} + SS_C + SS_{AC} + SS_{B(A)C}$$

$$SS_{ST2} = SS_A + SS_C + SS_{AC}$$

Procederemos a elaborar la tabla ANDEVA, calculando primeramente las interacciones AC, B(A)C y el ERROR.

$$SS_{AC} = SS_{ST2} - SS_A - SS_C = \underline{2.6406}$$

$$SS_{B(A)C} = SS_{ST1} - SS_{ST2} - SS_{B(A)} = \underline{7.88}$$

$$SS_E = SS_T - SS_{ST1} = \underline{15.79}$$

| Fuente | SS | GL | CM | F ₀ | F _{α,v1,v2} | DECISIÓN |
|----------|----------|----|---------------|----------------|----------------------|----------|
| A) Zona | 21.7539 | 2 | 10.8770 | 12.399 | 3.55 al 5% | SI |
| B) Oper. | 22.0533 | 6 | 3.6755 | 4.189 | 2.66 al 5% | SI |
| C) Proc. | 145.6044 | 1 | 145.6044 | 165.984 | 4.41 al 5% | SI |
| AC | 2.6406 | 2 | 1.3203 | 1.505 | 3.55 al 5% | NO |
| B(A)C | 7.8800 | 6 | 1.3133 | 1.497 | 2.66 al 5% | NO |
| Error | 15.7900 | 18 | 0.8772 | | | |
| Total | 215.7222 | 35 | | | | |

Los tres efectos, A, B(A) y C resultan ser significativos al 5%, las interacciones no lo son; al observar los resultados vemos que es fuertemente significativa la diferencia entre el procedimiento actual (2) y el procedimiento nuevo que está propuesto (1), notamos también que las zonas son significativas, la zona 1 es la que en general logra mejor promedio, pero la zona tres es la que mejora en mayor cantidad con el procedimiento propuesto.

4.5 Regresión Lineal

En muchas situaciones del área de calidad, existen una o más variables (**x**) que están relacionadas y resulta importante modelar y explorar las relaciones existentes entre ellas y el resultado que los diferentes valores generan de la variable respuesta (**y**), por ejemplo: El costo total de inventario por año (**y**) podemos sospechar que está relacionado con el nivel promedio de inventario (**x**) para ese mismo año y si existe 5 años de información podemos buscar la relación existente entre estas dos variables.

En nuestro curso estimado lector sólo nos abocaremos al estudio de relaciones lineales entre las variables (**x**'s) que las consideraremos independientes y la variable respuesta (**y**) que su comportamiento dependerá de las variables (**x**'s).

NOTACIÓN A UTILIZAR EN REGRESIÓN:

- n** Número de puntos experimentales (x's, y).
- m** Número de variables independientes que se tienen en Regresión lineal múltiple.
- i** Identificador del (i)ésimo punto experimental, que toma valores desde **1 a n**.
- j** Identificador de la (j)ésima variable independiente x, que toma valores desde **1 a m**.
- β₀** Parámetro del coeficiente constante del modelo.
- β_j** Parámetro del coeficiente de la (j)ésima variable x del modelo.
- b₀** Estimador del coeficiente constante del modelo de regresión.
- b₀'** Estimador del coeficiente constante "ajustado" del modelo regresión.
- b_j** Estimador del coeficiente de la (j)ésima variable x del modelo de regresión.
- X_{.j}** Media de la (j)ésima variable del modelo regresión.

- S_{jj}** Suma de cuadrados de la (j)ésima variable x.
- S_{yy}** Suma de cuadrados de la variable dependiente y.
- S_{kl}** Suma de cuadrados cruzados entre la variable (k) y la variable (l).
- y_i** Es el (i)ésimo dato de la variable dependiente y.
- ŷ_i** Es el (i)ésimo dato estimado por la regresión de la variable y.
- Y.** Media de la variable dependiente y.

La palabra ajustado o modificado la utilizaremos para describir el modelo de regresión manejado sólo con sus diferencias de cada dato con respecto a sus medias correspondientes.

REGRESIÓN LINEAL SIMPLE:

El modelo de regresión lineal simple es una línea recta, que sólo tiene una variable x independiente su modelo poblacional y el modelo de regresión estimado son:

| | |
|---|---|
| y_i = β₀ + β₁x_i + e_i | Modelo poblacional |
| y_i = β₀' + β₁(x_i - X.) + e_i | Modelo poblacional ajustado |
| ŷ_i = b₀ + b₁x_i | Modelo estimado por regresión |
| ŷ_i = b₀' + b₁(x_i - X.) | Modelo est. ajustado por regresión |

REGRESIÓN LINEAL MÚLTIPLE:

En este modelo se tienen (m) variables independientes x's las cuales se encuentran en relación lineal, y la variable dependiente (y), está relacionada a todas ellas por medio de las siguientes expresiones:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_m x_{im} + e_i$$

$$y_i = \beta_0' + \beta_1(x_{i1} - X_{.1}) + \beta_2(x_{i2} - X_{.2}) + \dots + \beta_j(x_{ij} - X_{.j}) + \dots + \beta_m(x_{im} - X_{.m}) + e_i$$

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_j x_{ij} + \dots + b_m x_{im}$$

$$\hat{y}_i = b_0' + b_1(x_{i1} - X_{.1}) + b_2(x_{i2} - X_{.2}) + \dots + b_j(x_{ij} - X_{.j}) + \dots + b_m(x_{im} - X_{.m})$$

En regresión lineal, podemos desglosar las diferencias provocadas en la variable dependiente (y) y obtener su correspondiente suma de cuadrados.

$$(y_i - Y.) = (\hat{y}_i - Y.) + (y_i - \hat{y}_i) \quad \text{por lo que obtenemos: } \mathbf{SS_T = SS_R + SS_E}$$

$$SS_T = \sum_{i=1}^n y_i^2 - \frac{T_Y^2}{n} = S_{YY} = \sum_{i=1}^n (y_i - Y.)^2$$

$$SS_R = \sum_{j=1}^m b_j S_{jY}$$

$$GL_T = n - 1 \qquad GL_R = m \qquad GL_E = n - m - 1$$

Uno de los métodos para determinar los coeficientes (b_j) es utilizando mínimos cuadrados, en la forma habitual que hicimos para el modelo de un factor, igualar a cero y obtener los valores de b_0' , b_1 . Para Regresión lineal simple obtenemos:

$$b_1 = \frac{S_{1Y}}{S_{11}} \quad b_0' = Y. \quad b_0 = Y. - b_1X.$$

Para regresión lineal múltiple, el procedimiento general es el mismo, aplicando vectores y matrices obtenemos la siguiente ecuación matricial:

$$Y = XB + E$$

Para el cálculo del vector B obtenemos la siguiente expresión:

$$B = (X^T X)^{-1} X^T Y$$

NOTA: El expositor desarrollará en clase el procedimiento general y ejemplos ilustrativos.

En la actualidad existen paquetería de Software la cual simplifica mucho la estimación de los coeficientes de la regresión.

EJEMPLO 4-5: La dureza del papel utilizado en la manufactura de cajas (y) está relacionado con la concentración en porcentaje de la pulpa original dura (x). Bajo condiciones controladas, la planta piloto genera 16 muestras, cada una desde diferentes lotes de pulpa, se hacen mediciones de porcentaje y dureza obteniendo la siguiente información.

| | | | | | | | | |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| y | 101.4 | 117.4 | 117.1 | 106.2 | 131.9 | 146.9 | 146.8 | 133.9 |
| x | 1.0 | 1.5 | 1.5 | 1.5 | 2.0 | 2.0 | 2.2 | 2.4 |
| y | 111.3 | 123.0 | 125.1 | 145.2 | 134.3 | 144.5 | 143.7 | 146.9 |
| x | 2.5 | 2.5 | 2.8 | 2.8 | 3.0 | 3.0 | 3.2 | 3.3 |

$$T_Y = 2075.6 \quad T_X = 37.2 \quad S_{YY} = 3650.81 \quad S_{XX} = 7.17 \quad S_{XY} = 112.2$$

Se desea buscar la relación lineal de estas dos variables, el coeficiente de determinación R^2 y el ANDEVA correspondiente.

$$b_1 = S_{XY}/S_{XX} = 15.6485 \quad b_0 = T_Y/n - b_1(T_X/n) = 93.3422$$

La relación lineal encontrada es: $\hat{y}_i = 93.3422 + 15.6485x_i$

$$SS_T = S_{YY} = 3650.81 \quad SS_R = b_1 S_{XY} = 1755.7657 \quad SS_E = 1895.0443$$

El coeficiente de determinación es la división de la suma de cuadrados debida a la regresión entre la suma de cuadrados totales.

$$R^2 = SS_R/SS_T = 1755.7657 / 3650.81 = 0.48092$$

El resultado anterior estimado lector, quiere decir que el 48% de las variaciones están justificadas por el modelo de regresión lineal, al hacer el ANDEVA se tomará la decisión si es significativa la relación lineal o no.

| F.V. | SS | GL | CM | F0 | F $_{\alpha, v1, v2}$ | Decisión |
|-----------|-----------|----|-----------|---------|-----------------------|----------|
| Regresión | 1755.7657 | 1 | 1755.7657 | 12.9711 | 4.60 al 5% | SI |
| Error | 1895.0443 | 14 | 135.3603 | | | |
| Total | 3650.8100 | 15 | | | | |

Aunque existe relación, el coeficiente de determinación nos guía a abocar nuestros esfuerzos a reducir la variabilidad aleatoria contenida en el error, o deberemos de buscar otra relación aunque no sea lineal.

4.6 Covarianza y correlación

Si tenemos dos variables aleatorias (x) y (y), la covarianza es una medida de la asociación lineal que pueda existir en dichas variables. El coeficiente de correlación tiene involucrada a la covarianza y describe el grado de asociación lineal que existe entre ambas variables, éste es adimensional y las expresiones poblacionales de ambos son las siguientes:

$$\text{Cov}(x,y) = \sigma_{xy} = E[(x - E(x))(y - E(y))] = E(x,y) - [E(x)E(y)] \quad \text{Covarianza } x,y$$

$$\rho = \frac{\text{Cov}(x,y)}{\sqrt{V(x)} \cdot \sqrt{V(y)}} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \quad \text{Coeficiente de correlación}$$

Si tenemos un tamaño de muestra (n), la expresión muestral para el coeficiente de correlación es la siguiente:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot \sum_{i=1}^n (y_i - \bar{Y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

El análisis de Covarianza es una técnica que resulta útil para mejorar la precisión de un experimento. Supongamos que en un experimento la variable de respuesta (y) está relacionada linealmente con la variable independiente (x) y existe la situación que no podemos controlar los valores de la variable (x), pero podemos medirla al mismo tiempo que la variable respuesta (y). En éste caso, decimos que la variable (x) es una **COVARIABLE** del problema o también algunos autores le llaman variable concomitante.

El modelo para el caso de un factor (A), que contiene (a) niveles de estudio, se tienen (n) réplicas o datos en cada nivel y la variable respuesta (y), tiene asociación lineal con la covariable (x) es el siguiente:

$$y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{X}_{..}) + e_{ij} \quad \text{donde } i = 1,2, \dots, a \quad \text{y} \quad j = 1,2, \dots, n$$

Observe estimado lector que el análisis de covarianza en este modelo es una combinación de los modelos lineales empleados en los análisis de varianza y regresión, a la covariable se le resta su media para que no modifique el valor de la gran media de los datos de (y), ya que el factor que se está estudiando es (A) representado en los valores de las y's.

SUPOSICIONES QUE CONSIDERA EL MODELO:

1. Los errores (e_{ij}) están normalmente distribuidos con media cero y varianza sigma cuadrada.
2. La pendiente beta es el coeficiente obtenido por regresión lineal y existe relación lineal, es decir, beta diferente a cero. Las pendientes a los diferentes (a) niveles del factor son iguales, es decir, el cambio de nivel no provoca modificación en el valor de la relación lineal entre (x) y (y).
3. La suma de las alfas valen cero.

NOTACIÓN Y FÓRMULAS PARA ANÁLISIS DE COVARIANZA:

NOTACIÓN GENÉRICA, TANTO PARA yy, xy O xx:

$$S = T + E$$

SUMAS DE CUADRADOS INVOLUCRADOS:

$$S_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - Y_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{T_y^2}{an}$$

$$S_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - X_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{T_x^2}{an}$$

$$S_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - X_{..})(y_{ij} - Y_{..}) = \sum_{i=1}^a \sum_{j=1}^n x_{ij}y_{ij} - \frac{T_x T_y}{an}$$

$$T_{yy} = \sum_{i=1}^a (Y_{i.} - Y_{..})^2 = \sum_{i=1}^a \frac{T_{yi.}^2}{n} - \frac{T_y^2}{an}$$

$$T_{xx} = \sum_{i=1}^a (X_{i.} - X_{..})^2 = \sum_{i=1}^a \frac{T_{xi.}^2}{n} - \frac{T_x^2}{an}$$

$$T_{xy} = \sum_{i=1}^a (X_{i.} - X_{..})(Y_{i.} - Y_{..}) = \sum_{i=1}^a \frac{T_{xi.} T_{yi.}}{n} - \frac{T_x T_y}{an}$$

$$E_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - Y_{i\cdot})^2 = S_{yy} - T_{yy}$$

$$E_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - X_{i\cdot})^2 = S_{xx} - T_{xx}$$

$$E_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - X_{i\cdot})(y_{ij} - Y_{i\cdot}) = S_{xy} - T_{xy}$$

Para el modelo de esta sección, los estimadores de mínimos cuadrados de la media, alfas (i) y beta son: Para la media $Y_{i\cdot}$, para la $\alpha_i = Y_{i\cdot} - Y_{..} - b(X_{i\cdot} - X_{..})$ y, para $b = E_{xy}/E_{xx}$.

La suma de cuadrados del error del modelo es:

$$SS_E = E_{yy} - (E_{xy})^2 / E_{xx} \quad \text{con } a(n - 1) - 1 \text{ GL.} \quad CM_E = SS_E / [a(n - 1) - 1]$$

Si no existiera efecto producido por los niveles (i), entonces se reduce el modelo y los estimadores por mínimos cuadrados son para la media, $Y_{..}$, para beta es: $b = S_{xy}/S_{xx}$.

La suma de cuadrados (prima) para esta condición es:

$$SS'_E = S_{yy} - (S_{xy})^2 / S_{xx} \quad \text{con } an - 2 \text{ GL.}$$

La suma de cuadrados del error es menor que la suma de cuadrados del error (prima) porque los parámetros alfa adicionales se encuentran en el modelo y la diferencia de ambos es una reducción de la suma de cuadrados provocada por las alfas, por lo tanto su diferencia tiene (a - 1) grados de libertad.

$(SS'_E - SS_E)$ con $(a - 1)$ GL. debidos a los niveles alfa.

Para probar entonces las alfas tendremos el estadístico F_0 de la siguiente forma:

$$F_0 = \frac{(SS'_E - SS_E) / (a - 1)}{CM_E}$$

Note estimado lector que si no hubiere **COVARIABLE** tendríamos lo siguiente:

$$S_{xy} = S_{xx} = E_{xy} = E_{xx} = 0$$

En ese caso la suma de cuadrados del error simplemente sería E_{yy} . La suma de cuadrados de los niveles sería $T_{yy} = S_{yy} - E_{yy}$.

EJEMPLO 4-6:

Se utilizan 3 máquinas para producir fibras de solo un filamento para una compañía textil, el investigador del proceso está interesado en determinar si existe diferencia en la resistencia a la ruptura de la fibra producida por las tres máquinas. Sin embargo, la resistencia de una fibra depende del grosor de la misma, siendo generalmente más resistentes las fibras de mayor grosor. Se selecciona una muestra al azar de 5 fragmentos de fibra por cada máquina, se miden la resistencia de cada fibra (y) y el grosor correspondiente (x), los datos obtenidos son (y) es resistencia en libras y (x) es diámetro en milésimas de pulgada.

| | Máq. 1 | | Máq. 2 | | Máq. 3 | |
|-----------------------|--------|-----|--------|-----|--------|-----|
| | y | x | y | x | y | x |
| | 36 | 20 | 40 | 22 | 35 | 21 |
| | 41 | 25 | 48 | 28 | 37 | 23 |
| | 39 | 24 | 39 | 22 | 42 | 26 |
| | 42 | 25 | 45 | 30 | 34 | 21 |
| | 49 | 32 | 44 | 28 | 32 | 15 |
| T_{i.} | 207 | 126 | 216 | 130 | 180 | 106 |

T_y = 603

T_x = 362

a = 3

n = 5

S_{yy} = 346.40

S_{xx} = 261.73

S_{xy} = 282.60

T_{yy} = 140.40

T_{xx} = 66.13

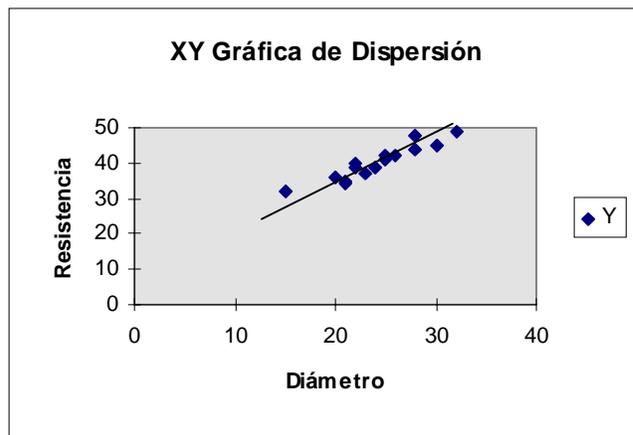
T_{xy} = 96.00

E_{yy} = 206.00

E_{xx} = 195.60

E_{xy} = 186.60

Si obtenemos su diagrama de dispersión x,y observamos que si existe una relación lineal dada por el diámetro a la resistencia a la ruptura.



El problema puede verse desde dos enfoques, al ser consistente generan el mismo resultado estadístico; el primero es verlo como "Análisis de varianza ajustado", el segundo es verlo como "Análisis de Covarianza de un factor con una covariable".

Por la sección en que estamos nos interesa más el segundo, pero lo veremos en ambos enfoques:

ANÁLISIS DE VARIANZA "AJUSTADO" para el Error:

| F. de V. | SS | GL | CM | Fo |
|-----------------|--------------------------------------|----------------|--------------------------------|-------------|
| Error Regresión | $SS_E' = S_{yy} - (S_{xy})^2/S_{xx}$ | $an - 2$ | | |
| Factor ó Trat. | $SS_E' - SS_E$ | $a - 1$ | $CM_F = (SS_E' - SS_E)/(a-1)$ | CM_F/CM_E |
| Error | $SS_E = E_{yy} - (E_{xy})^2/E_{xx}$ | $a(n - 1) - 1$ | $CM_E = \frac{SS_E}{a(n-1)-1}$ | |
| Total | S_{yy} | $an - 1$ | | |

ANÁLISIS DE COVARIANZA DE UN FACTOR CON UNA COVARIABLE:

| | | Suma de cuadrados y Productos | | | |
|----------|----------|-------------------------------|----------|----------|--------------------------------------|
| F. de V. | GL | x | xy | y | y(ajustada) |
| Factor | $a - 1$ | T_{xx} | T_{xy} | T_{yy} | |
| Error | $a(n-1)$ | E_{xx} | E_{xy} | E_{yy} | $SS_E = E_{yy} - (E_{xy})^2/E_{xx}$ |
| Total | $an - 1$ | S_{xx} | S_{xy} | S_{yy} | $SS_E' = S_{yy} - (S_{xy})^2/S_{xx}$ |

Y ahora el ajuste de la regresión para el Factor:

| F. de V. | SS | GL | CM | Fo |
|-----------------|----------------|----------------|--------|-------------|
| Factor (ajust.) | $SS_E' - SS_E$ | $a - 1$ | CM_F | CM_F/CM_E |
| Error | SS_E | $a(n - 1) - 1$ | CM_E | |
| Total | SS_E' | $an - 2$ | | |

Para el ejemplo 4-6, los valores son:

$SS_E' = 41.2659$ con 13 grados de libertad. $SS_E = 27.9859$ con 11 grados de libertad.
 $CM_F = 13.28/2 = 6.64$ con 2 grados de libertad. $CM_E = 2.5442$ con 11 grados de libertad.

$Fo = 2.6099$ con 2 GL en numerador y 11 GL en denominador. En la tabla al 5% se compara contra 3.98 por lo que no resulta significativa la diferencia de los niveles de (y). Lo cual implica que se reduce el modelo simplemente a regresión por mínimos cuadrados y los estimadores son: $Y.. = 40.2$ α_i aprox. cero $b = S_{xy}/S_{xx} = 1.0797$.

No es el caso de este resultado, pero, si hubiera sido significativo, entonces adicionalmente a la regresión hubiera existido el factor máquinas presente y tendríamos que haber calculado los coeficientes del modelo en la siguiente forma:

$$Y.. = 40.2 \quad \alpha_i = Y_{i.} - Y.. - b(X_{i.} - X..) \quad b = E_{xy}/E_{xx}$$

Para el ejemplo 4-6, la recta de regresión es: $\hat{y}_i = 14.1426 + 1.0797x_i$
 Con un **Coefficiente de Determinación de:** $R^2 = 0.8809$
 El **Coefficiente global de Regresión es:** $R = 0.9385$

5. Contribuciones de Taguchi

En todo lo anterior en estos apuntes, se ha dado importancia a utilizar experimentos diseñados para el mejoramiento de productos, servicios y procesos. Una meta importante del mejoramiento de la Calidad, es incorporar la **Calidad de Diseño** de un servicio o producto desde el proceso que éste se elabora. Los experimentos diseñados estadísticamente son un elemento importante para lograr mejoramientos perdurables de Calidad.

El Dr. Genichi Taguchi, a finales de los setentas introdujo su método para usar el Diseño Experimental en:

1. **Diseño de productos o procesos de modo que sean robustos ante las condiciones ambientales.**
2. **Diseño y/o Desarrollo de productos que sean robustos a la variación de componentes.**
3. **Minimización de las variaciones respecto a un valor objetivo.**

FILOSOFÍA DE TAGUCHI:

Ésta es una filosofía de la Ingeniería de Calidad que es ampliamente aplicable. Él considera tres etapas en el desarrollo de un producto o proceso, estas son: **Diseño del Sistema, Diseño de Parámetros y Diseño de Tolerancias.**

Teniendo en consideración que la interpretación de Calidad en las cosas corresponde al correcto entendimiento de que Calidad es: **LA MEJOR ADECUACIÓN DE USO DE LAS COSAS, PARA LO CUAL FUERON CREADAS.**

Taguchi recomienda el empleo de métodos estadísticos para auxiliar el mejoramiento de la Calidad, en particular durante el diseño de parámetros y el diseño de tolerancias. Pueden emplearse métodos de Diseño Experimental para hallar un mejor diseño del producto o del proceso, donde por “**mejor**” quiere decirse un producto o proceso que es ROBUSTO o insensible a factores incontrolables que influirán en ese producto o proceso una vez que se encuentre en funcionamiento normal.

La noción de diseño robusto o consistente no es nueva. Los Ingenieros siempre han tratado de diseñar productos que funcionen bien en condiciones incontrolables. Por ejemplo, los aviones de transporte comercial vuela tan bien en medio de una tormenta eléctrica como en un día despejado. Taguchi merece el crédito por haber observado que el diseño experimental puede utilizarse como una parte formal del proceso técnico (o de Ingeniería) para ayudar a alcanzar el objetivo de consistencia.

En general las etapas que plantea Ingeniería de Calidad son:

DISEÑO DEL SISTEMA: Denota el desarrollo de Diseños Prototipos Básicos, los cuales se configuran a las funciones deseadas y requeridas del producto, con mínimas desviaciones de los valores objetivos configurados. Ésta incluye la selección de materiales, partes, componentes y el sistema de ensamblaje, como ejemplo de la idea de prototipo, podríamos pensar en un circuito (fuente de poder de computadora) para convertir 115 Volts corriente alterna a +5, +12, -5 y -12 Volts corriente continua que alimentará las partes electrónicas y mecánicas de una computadora.

El circuito anterior requiere una buena investigación técnica que genere el mejor circuito, que específicamente sea relevante el diseño obtenido, ya que cualquier variabilidad del voltaje nominal de +5 y -5 provocará reducción de la vida esperada (ciclo de vida) en los componentes de la computadora y por ende en una o en otra forma creará costos adicionales e incomodidades para el Cliente final, distribuidor y fabricante.

DISEÑO DE PARÁMETROS: En ésta etapa, debemos encontrar los niveles (valores) de los factores confortables, para nuestro producto, éstos deben ser seleccionados tales que minimicen los efectos de factores de ruido (noise) sobre las características funcionales del producto.

DISEÑO DE TOLERANCIAS: En ésta etapa deberemos encontrar los valores que permitan hacer adecuado el funcionamiento flexible y real, la propuesta en que se basa Taguchi es la inserción al manejo estadístico de los datos, de una **Función de Pérdida asociada al nivel de Calidad** exigido por los atributos y condiciones a cumplir, obviamente dicha función de pérdida involucra costos, que de una o de otra forma deberán ser pagados por el fabricante, distribuidor o cliente final (elementos de la sociedad).

Es necesario definir rangos permitibles de desviación del parámetro; a rangos muy estrechos de variación provocará un costeo mayor del producto que se reflejará en incrementos de costos de manufactura o fabricación. En cambio, un amplio rango de variación en la función del producto comparado contra el valor específico objetivo, generará una reducción de la adecuación al uso del producto y por ende, costos de insatisfacción, mantenimiento y decrecimiento fuerte en la aceptación del cliente final. El Diseño de Tolerancias determina las mejores, tal que minimice el costo del producto.

A sido grandemente utilizada estas tres etapas en muchas industrias para implementar la calidad de los productos. Las actividades de Control de Calidad en las fases de Planeación de productos, Diseño e Ingeniería de producción se han considerado como actividades de Control de Calidad “fuera de línea de producción” o Ingeniería de Calidad, en cambio las actividades de Control de Calidad durante la producción real han sido referenciadas como actividades de Control de Calidad “en línea de producción”.

Taguchi & Wu (1979) resumen estas actividades para una típica empresa manufacturera en la siguiente tabla, en donde, un asterisco en las columnas indican que el ruido externo, ruido interno y la cantidad de unidades de variación (tolerancias) pueden ser controlados en cada una de las fases del ciclo de vida del producto. Un signo más (+) indica que no es preferible el control del ruido, es decir, debe buscarse desde el diseño del sistema su eliminación y un doble signo más (++) indica que es imposible el control del ruido en esa fase.

ACTIVIDADES FUNCIONALES DE CONTROL DE CALIDAD:

| Act. de Control de Calidad | Fase del Producto | Fases | Ruido Externo | Ruido Interno | Cant. de Unidades de Variación |
|----------------------------|--------------------------|-----------------------|---------------|---------------|--------------------------------|
| “Fuera de Línea” | Diseño del Producto | Diseño del Sistema | * | * | * |
| | | Diseño de Parámetros | * | * | * |
| | | Diseño de Tolerancias | + | * | * |
| | Ingeniería de Producción | Diseño del Sistema | ++ | ++ | * |
| | | Diseño de Parámetros | ++ | ++ | * |
| | | Diseño de Tolerancias | ++ | ++ | * |
| “En Línea” | Operación de Producción | Control del Proceso | ++ | ++ | * |
| | | Retroalimentación | ++ | ++ | * |
| | | Inspección, etc. | ++ | ++ | * |

5.1 Función de pérdida de Calidad

Consideremos dos fábricas de una misma organización, la fábrica (A) y la fábrica (B), ambas elaboran televisores con el mismo proceso productivo y el mismo diseño técnico, ambas fábricas tienen las mismas especificaciones y utilizan las mismas materias primas para la fabricación, es de esperarse por lo tanto que ambas fábricas pudieran obtener también los mismos niveles de calidad y sea indistinta la preferencia que los clientes finales puedan tener por los televisores considerando que tienen el mismo valor. Lo anterior es eminentemente lógico pero, en el párrafo anterior no se ha considerado la forma en que los trabajadores de ambas fábricas logran cumplir con las especificaciones de alguna característica funcional de calidad.

Supongamos que en la fábrica (A), los trabajadores están entrenados conceptualmente para entender que cuando se da un valor (parámetro de diseño) característico que se debe cumplir en el equipo y se permite una tolerancia a dicho valor, están conscientes de que el ajuste deberá ser lo más cercano al parámetro, es decir, es más importante estar cerca del valor característico no importando la amplitud o estreches de la tolerancia permitida.

En cambio, en la fábrica (B), los trabajadores están entrenados para cumplir con las normas y especificaciones de fabricación, para estos trabajadores es más importante cumplir “dentro de tolerancias” los requerimientos de producción y menos importante estar cerca al valor característico de diseño. Podemos adicionar que en la fábrica (B) todos los televisores que salen de ella se encuentran “dentro de especificaciones”, en cambio algunos pocos televisores de la fábrica (A) no lo están.

La pregunta inductiva aquí estimado lector es: ¿Qué fábrica tiene mejor Calidad con respecto al parámetro de diseño?, para poder contestar esta pregunta debemos tener información sobre “esa pequeña cantidad de televisores de la fábrica (A) que no están dentro de especificaciones” y en segundo lugar es “qué tan costoso está siendo para la fábrica (B) sacar todos sus televisores dentro de las especificaciones”.

Podrá parecerle estimado lector que lo planteado anteriormente sea mera especulación mental, **NO LO ES**, la estrategia de los trabajadores de la fábrica (A) conlleva una actitud de **EXACTITUD Y PRECISIÓN**, en cambio la estrategia de los trabajadores de la fábrica (B) conlleva una actitud de **EXACTITUD PERO IMPRECISIÓN**.

Un ejemplo de este tipo fue publicado por la revista Ashi el 17 de abril de 1979 en Tokio, con una empresa internacional que tiene una planta en Estados Unidos y la otra en Japón y fabrica televisores.

Para ejemplificar esta situación consideraremos nuestras dos fábricas (A) y (B), compararemos la Calidad de concentración de color producido en ambas fábricas, los límites totales de tolerancia es **10 (más, menos 5** de ambos lados) ambas fábricas tienen distribuciones de concentración de color centradas al valor del parámetro de diseño **m = 30** la fábrica (A) tiene una distribución normal con $\sigma = 10/6 = 1.6666$ y la fábrica (B) tiene una distribución uniforme con $\sigma^2 = 100/12$ y su desviación estándar es: $\sigma = 2.8868$. (ver figura No. 5-1), note estimado lector que la distribución de la fábrica (A) concentra mayor número de televisores al centro de la característica funcional de calidad (de 28 a 32 el 77% aprox. y de 29 a 31 el 45.15% aprox.), en cambio en la distribución de la fábrica (B) concentra menos número de televisores al centro (de 28 a 32 el 40% y de 29 a 31 el 20%), puede comprobarlo en la tabla de datos de la sección 5.1.2 que está posteriormente.

Taguchi plantea inicialmente, relaciones cuadráticas para estimar la **función de pérdida de calidad**, por no hacer el ajuste desde un principio en forma **exacta y precisa** en el parámetro funcional de la característica de calidad, desglosa tres casos genéricos, estos son:

1. **VALOR NOMINAL ES MEJOR:** (The nominal the best, N type) El valor nominal del parámetro funcional de calidad es referenciado al valor cero de costo de reparación ya que se considera como valor de diseño y está interrelacionado con las condiciones de calidad generadas en el diseño de parámetros de Ingeniería de Calidad (tipo N).
2. **VALOR MÁS PEQUEÑO ES EL MEJOR:** (The smaller the better, S type) El valor del parámetro funcional de calidad generado por el diseño de parámetros de Ingeniería de Calidad es cero ya que la variable de calidad manejada es no deseada.
3. **VALOR MÁS GRANDE ES EL MEJOR:** (The larger the better, L type) El valor del parámetro funcional de calidad generado por el diseño de parámetros de Ingeniería de Calidad es teóricamente infinito ya que la variable de calidad manejada es totalmente deseada.

En estos tres casos genéricos, Taguchi plantea la función de pérdida de calidad con las siguientes expresiones, siendo **L(x)** la función y tiene unidades monetarias por cada producto desajustado, siendo **k** la constante de proporcionalidad teniendo unidades tales que la función tenga unidades monetarias por cada producto desajustado, siendo **m** el valor del parámetro característico de calidad a cumplir:

NOMINAL ES MEJOR (TIPO N): $L(x) = k(x - m)^2$

PEQUEÑO ES MEJOR (TIPO S): $L(x) = kx^2$

GRANDE ES MEJOR (TIPO L): $L(x) = \frac{k}{x^2}$

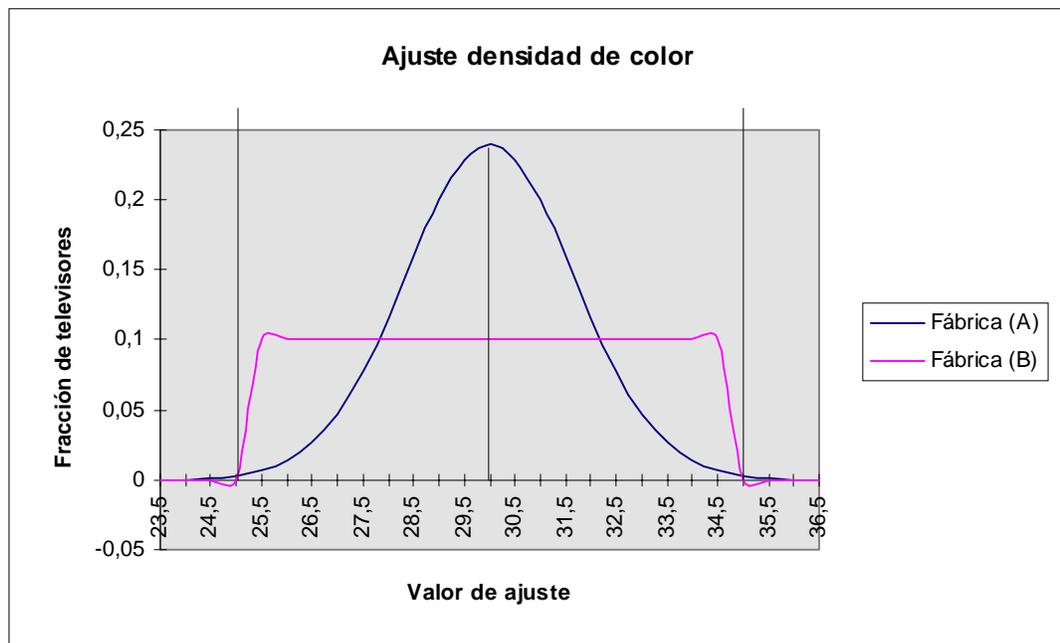


Fig. No. 5-1

Para los casos (1) y (2) de la función de pérdida de calidad es fácil demostrar que el valor esperado de la función de pérdida aplicando la ecuación de expectación de la sección 2.3

$$E(L(x)) = k(\sigma^2 + D^2)$$

$$D^2 = (\mu - m)^2$$

En donde **D** representa la diferencia o sesgo cuando la distribución no se encuentre centrada al parámetro característico de calidad.

En cambio, para el caso (3) de la función de pérdida de calidad (Mayor es lo mejor), podemos hacer una sustitución de variable $y = 1/x$, siendo (x) la variable aleatoria en estudio con valor de objetivo infinito, manejamos la variable (y) aleatoria con una función de pérdida equivalente a "Menor es lo mejor" con valor (y) mayor o igual a cero.

$$E(L(y)) = k_y(\sigma_y^2 + D_y^2)$$

5.1.1 Determinación de la constante de la función de pérdida

Para obtener el valor de la constante de proporcionalidad en la función de pérdida de calidad, debemos tener la información verificada del costo generado por producto defectuoso y/o el costo de reproceso a la tolerancia unilateral (casos (2) y (3)) o bilateral (caso (1)) y aplicando ésta condición numérica a la función de pérdida, despejamos la constante:

NOMINAL ES MEJOR: Si la tolerancia de los dos lados es: $\delta = (LSE - m) = (m - LIE)$ y el costo de pérdida en producto defectuoso o reproceso es (A) en las especificaciones, entonces el valor de (k) es: $k = A/\delta^2$

MENOR ES MEJOR: Si el límite superior de especificación (LSE) tiene un costo de (A), entonces el valor de (k) es: $k = A/(LSE)^2$

MAYOR ES MEJOR: Si el límite inferior de especificación (LIE) tiene un costo de (A), entonces el valor de (k) es: $k = A/(LIE)^2$

Para nuestro ejemplo de las fábricas (A) y (B), considere que se efectuó el estudio de costos y un televisor que haya sido mal ajustado en 25 o 35 (tolerancia unilateral $\delta = 5$), cuesta a la sociedad (fabricante, distribuidor y/o cliente), **A = \$ 100.00** (ver figura No. 5-2), calcularemos la constante de la función de pérdida.

$$k = \frac{A}{\delta^2} = \frac{100}{5^2} = 4 \frac{\$}{\text{unidades}^2}$$

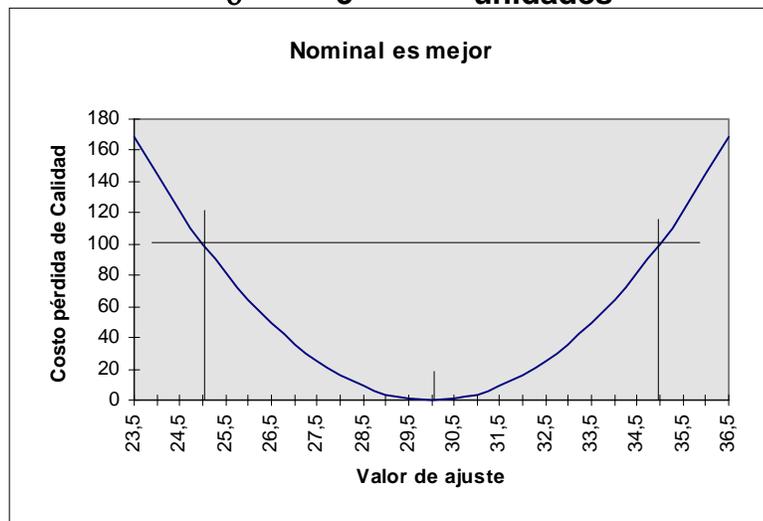


Fig. 5-2

5.1.2 Estimación de la pérdida esperada

Ya habiendo calculado la constante de la función de pérdida, podemos proceder a la estimación de la pérdida esperada por no hacer un **ajuste exacto y preciso** desde la producción inicial.

Para las fábricas (A) y (B) aplicaremos la misma función de pérdida para nuestro ejemplo, pero en la práctica pueden diferir algo, a cada fábrica se le aplicaría su propia función de pérdida.

$$L(x) = k(x - m)^2 \quad \text{en donde } k = 4 \text{ y } m = 30$$

Para estimar la pérdida esperada en el lote completo de producción, podemos utilizar la expresión del valor esperado calculado en la sección 5.1 si conocemos su varianza, o bien podemos estimar el valor de la varianza de la distribución por medio de un histograma de datos, si la producción es continua primero es preferible evaluar la distribución de probabilidad que sigue la producción y evaluamos su varianza, para directamente estimar la pérdida esperada total de producción por cada parámetro característico de calidad que esté involucrado en el proceso con la ecuación de valor esperado de la sección 5.1.

En nuestro ejemplo, tenemos las varianzas predeterminadas en la sección 5.1 y tienen los siguientes valores: **Fábrica (A)**, su valor es $\sigma^2 = 2.7777$ ($\sigma = 1.6666$) y para la **Fábrica (B)**, su valor es $\sigma^2 = 8.3333$ ($\sigma = 2.8868$).

Podemos calcular sus capacidades de proceso Cp de ambas fábricas, pero la (B) no sigue una normal, por lo que una comparación de Cp's es inconsistente ya que son diferentes distribuciones.

$$\text{Para (A): } C_p = 10/(6\sigma) = 1 \quad \text{Para (B): } C_p = 10/(6\sigma) = 0.5773$$

Estimaremos el valor esperado de la función de pérdida: En donde la $D = 0$.

$$\text{(A): } E(L(x)) = k(\sigma^2 + D^2) = \$ 11.11$$

$$\text{(B): } E(L(x)) = k(\sigma^2 + D^2) = \$ 33.33$$

Ahora bien, si no tuviéramos las varianzas predeterminadas como en este ejemplo, podemos evaluar las varianzas desde histograma de frecuencias relativas y aplicar la fórmula de expectación de la sección 2.3 para condiciones discretas.

$$\sigma^2 = \sum_{i=1}^c (x_i - \mu)^2 P(x_i) \quad \text{en donde } (c) \text{ es el número de clases}$$

El valor de las varianzas podrá diferir un poco ya que la creación del histograma presupone todos los valores de una clasificación como su valor medio o característico, calcularemos su varianza de (A) y (B) y sus valores esperados de la función de pérdida.

| X | VM | L(x) | Prob. A | Prob. B | P. Esp. A | P. Esp. B |
|--------------|-------------|----------|-----------|----------|-----------------|-----------------|
| 23 a 24 | 23.5 | 169.00 | ,00015906 | 0 | 0,02688114 | 0,00 |
| 24 a 25 | 24.5 | 121.00 | ,00119038 | 0 | 0,14403598 | 0,00 |
| 25 a 26 | 25.5 | 81.00 | ,00684594 | 0,1 | 0,55452114 | 8,10 |
| 26 a 27 | 26.5 | 49.00 | ,02772920 | 0,1 | 1,35873080 | 4,90 |
| 27 a 28 | 27.5 | 25.00 | ,07913583 | 0,1 | 1,97839575 | 2,50 |
| 28 a 29 | 28.5 | 9.00 | ,15918466 | 0,1 | 1,43266194 | 0,90 |
| 29 a 30 | 29.5 | 1.00 | ,22575493 | 0,1 | 0,22575493 | 0,10 |
| 30 a 31 | 30.5 | 1.00 | ,22575493 | 0,1 | 0,22575493 | 0,10 |
| 31 a 32 | 31.5 | 9.00 | ,15918466 | 0,1 | 1,43266194 | 0,90 |
| 32 a 33 | 32.5 | 25.00 | ,07913583 | 0,1 | 1,97839575 | 2,50 |
| 33 a 34 | 33.5 | 49.00 | ,02772920 | 0,1 | 1,35873080 | 4,90 |
| 34 a 35 | 34.5 | 81.00 | ,00684594 | 0,1 | 0,55452114 | 8,10 |
| 35 a 36 | 35.5 | 121.00 | ,00119038 | 0 | 0,14403598 | 0,00 |
| 36 a 37 | 36.5 | 169.00 | ,00015906 | 0 | 0,02688114 | 0,00 |
| Sumas | Esp. | = | 1 | 1 | \$ 11,44 | \$ 33,00 |

Habr  notado estimado lector, que para las dos f bricas (A) y (B) existen peque as diferencias con los valores de \$ 11,11 y con \$ 33.33 puesto que se est  suponiendo los valores medios de cada clasificaci n, los valores de las varianzas son:

(A): $\sigma^2 = 2.86049084$

(B): $\sigma^2 = 8.25$

Es importante notar que la f brica (A) s lo est n fuera de especificaciones el 0.27% de los televisores que fabrica y la f brica (B) por cumplir ese 0.27% tiene triplicado el costo esperado de ajuste de la densidad de color. **CONCLUSI N:** La f brica (A) responde mejor a las expectativas de adecuaci n de uso (concepto de Calidad).

5.2 Resoluci n de problemas

Para poder atacar los problemas con la filosof a de Taguchi, debemos tener en cuenta el siguiente procedimiento y aplicarlo correctamente.

1. **ESTUDIO Y DEFINICI N DEL PROBLEMA**
2. **DETERMINACI N Y DEFINICI N DEL OBJETIVO BUSCADO**
3. **LLUVIA DE IDEAS PARA POSIBLES SOLUCIONES**
4. **DISE O DEL EXPERIMENTO**
5. **GENERACI N Y CONDUCCI N DEL EXPERIMENTO PARA OBTENER DATOS**
6. **AN LISIS DE LA INFORMACI N OBTENIDA**
7. **INTERPRETACI N DE RESULTADOS**
8. **EJECUTAR NUEVO EXPERIMENTO Y/O CORRIDA CONFIRMATORIA**
9. **ELABORAR REPORTE COMPARATIVO**

Este procedimiento es general y puede aplicarse a cualquier proceso con caracter sticas cualitativas y/o cuantitativas, en el m todo de Taguchi y principalmente en el paso 4 (Dise o del Experimento) se utilizan arreglos ortogonales predeterminados que tienen la siguiente convenci n:

$L_e (a^c)$

En donde (e) es el n mero de corridas experimentales, (a) es el n mero de niveles para cada efecto o factor (se pueden incluir algunas interacciones de inter s) y (c) es el n mero de columnas en el arreglo ortogonal.

Se han desarrollado dos series de arreglos ortogonales, la que considera dos niveles por efecto o factor (**a = 2**), y la que considera tres niveles por efecto o factor (**a = 3**). La serie **a=2**, proviene de un sistema **2^k** y la serie **a=3** proviene de un sistema **3^k**.

Los arreglos **L₄**, **L₈**, **L₁₂**, **L₁₆**, **L₂₀**, **L₂₄**, **L₂₈** y **L₃₂** son de dos niveles por efecto o factor; los arreglos **L₉** y **L₂₇** son de tres niveles por efecto o factor, y el arreglo **L₁₈** es un arreglo ortogonal especial combinado (una columna tiene dos niveles y siete columnas tienen tres niveles), se muestran a continuaci n:

L₄ (2³) **L₈ (2⁷)** **L₉ (3⁴) NO USARLO EN INTERACCIONES**

| L ₄ | 1 | 2 | 3 | | L ₈ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | L ₉ | 1 | 2 | 3 | 4 |
|----------------|---|---|---|--|----------------|---|---|---|---|---|---|---|--|----------------|---|---|---|---|
| 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 | | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | | 2 | 1 | 2 | 2 | 2 |
| 3 | 2 | 1 | 2 | | 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | | 3 | 1 | 3 | 3 | 3 |
| 4 | 2 | 2 | 1 | | 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | | 4 | 2 | 1 | 2 | 3 |
| | | | | | 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | | 5 | 2 | 2 | 3 | 1 |
| | | | | | 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | | 6 | 2 | 3 | 1 | 2 |
| | | | | | 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | | 7 | 3 | 1 | 3 | 2 |
| | | | | | 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | | 8 | 3 | 2 | 1 | 3 |
| | | | | | | | | | | | | | | 9 | 3 | 3 | 2 | 1 |

L₁₂(2¹¹) NO USARLO EN INTERACCIONES

| L ₁₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 4 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |
| 5 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| 6 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 7 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 |
| 8 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 |
| 9 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| 10 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 |
| 11 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 12 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 |

L₁₆(2¹⁵)

| L ₁₆ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| 5 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 6 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 7 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 8 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| 9 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 10 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| 11 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 12 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 13 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 14 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 |
| 15 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 16 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |

L₁₈(2¹ x 3⁷) NO USARLO EN INTERACCIONES

| L ₁₈ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------------|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 3 |
| 5 | 1 | 2 | 2 | 2 | 3 | 3 | 1 | 1 |
| 6 | 1 | 2 | 3 | 3 | 1 | 1 | 2 | 2 |
| 7 | 1 | 3 | 1 | 2 | 1 | 3 | 2 | 3 |
| 8 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 1 |
| 9 | 1 | 3 | 3 | 1 | 3 | 2 | 1 | 2 |
| 10 | 2 | 1 | 1 | 3 | 3 | 2 | 2 | 1 |
| 11 | 2 | 1 | 2 | 1 | 1 | 3 | 3 | 2 |
| 12 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 |
| 13 | 2 | 2 | 1 | 2 | 3 | 1 | 3 | 2 |
| 14 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 3 |
| 15 | 2 | 2 | 3 | 1 | 2 | 3 | 2 | 1 |
| 16 | 2 | 3 | 1 | 3 | 2 | 3 | 1 | 2 |
| 17 | 2 | 3 | 2 | 1 | 3 | 1 | 2 | 3 |
| 18 | 2 | 3 | 3 | 2 | 1 | 2 | 2 | 1 |

L₂₀ (2¹⁹) NO USARLO EN INTERACCIONES

| L ₂₀ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 |
| 5 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 |
| 6 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |
| 7 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 |
| 8 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 |
| 9 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 |
| 10 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 |
| 11 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 |
| 12 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| 13 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| 14 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |
| 15 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| 16 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 |
| 17 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 |
| 18 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 |
| 19 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 |
| 20 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 |

L₂₄ (2²³) NO USARLO EN INTERACCIONES

| L ₂₄ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 5 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 6 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 |
| 7 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| 8 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| 9 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 |
| 10 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 |
| 11 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 |
| 12 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 |
| 13 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 |
| 14 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 15 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| 16 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 |
| 17 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| 18 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 |
| 19 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 20 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 |
| 21 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| 22 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 23 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 2 |
| 24 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 |

L₂₇ (3¹³)

| L ₂₇ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 |
| 5 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 1 |
| 6 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 1 | 2 | 2 | 2 |
| 7 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 3 | 3 | 3 | 2 | 2 | 2 |
| 8 | 1 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 3 | 3 |
| 9 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 1 |
| 10 | 2 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 11 | 2 | 1 | 2 | 3 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 |
| 12 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 |
| 13 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 2 | 3 | 1 | 3 | 1 | 2 |
| 14 | 2 | 2 | 3 | 1 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 3 |
| 15 | 2 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 1 |
| 16 | 2 | 3 | 1 | 2 | 1 | 2 | 3 | 3 | 1 | 2 | 2 | 3 | 1 |
| 17 | 2 | 3 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 3 | 1 | 2 |
| 18 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 3 |
| 19 | 3 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 |
| 20 | 3 | 1 | 3 | 2 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 | 3 |
| 21 | 3 | 1 | 3 | 2 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 2 | 1 |
| 22 | 3 | 2 | 1 | 3 | 1 | 3 | 2 | 2 | 1 | 3 | 3 | 2 | 1 |
| 23 | 3 | 2 | 1 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 1 | 3 | 2 |
| 24 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 2 | 1 | 3 |
| 25 | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 3 | 2 | 1 | 2 | 1 | 3 |
| 26 | 3 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 | 3 | 2 | 1 |
| 27 | 3 | 3 | 2 | 1 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 |

L₂₈ (2²⁷) NO USARLO EN INTERACCIONES

| L ₂₈ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 4 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| 5 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 |
| 6 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 |
| 7 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 |
| 8 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 |
| 9 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 10 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 |
| 11 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 |
| 12 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 |
| 13 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 |
| 14 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| 15 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| 16 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 2 |
| 17 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 |
| 18 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 19 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 |
| 20 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 |
| 21 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 22 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 23 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 |
| 24 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 |
| 25 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 26 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| 27 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 |
| 28 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |

Estos arreglos ortogonales no fueron específicamente desarrollados por Taguchi, por ejemplo, el L_8 es un factorial fraccionario del tipo 2_{III}^{7-4} así algunos otros son:

L_9 es factorial fraccionario 3_{III}^{4-2}

L_{12} es Diseñ ode Plackett Burman

L_{16} es factorial fraccionario 2_{III}^{15-11}

Cada una de las columnas de los arreglos ortogonales, se asocia con un efecto en estudio (factor o interacción), para asignar correctamente las columnas que deben corresponder a cada interacción de interés en el estudio es necesario colocar primero los factores en estudio e inmediatamente después colocar su interacción de interés conforme a la siguiente tabla estándar de interacciones para sistemas 2^C . Ésta es válida para L_4 , L_8 y L_{16} .

NOTA: Los efectos parcialmente confundidos se han omitido.

| Efecto | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|--------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | | C | B | E | D | G | F | I | H | K | J | M | L | O | N |
| B | C | | A | F | G | D | E | J | K | H | I | N | O | L | M |
| C | B | A | | G | F | E | D | K | J | I | H | O | N | M | L |
| D | E | F | G | | A | B | C | L | M | N | O | H | I | J | K |
| E | D | G | F | A | | C | B | M | L | O | N | I | H | K | J |
| F | G | D | E | B | C | | A | N | O | L | M | J | K | H | I |
| G | F | E | D | C | B | A | | O | N | M | L | K | J | I | H |
| H | I | J | K | L | M | N | O | | A | B | C | D | E | F | G |
| I | H | K | J | M | L | O | N | A | | C | B | E | D | G | F |
| J | K | H | I | N | O | L | M | B | C | | A | F | G | D | E |
| K | J | I | H | O | N | M | L | C | B | A | | G | F | E | D |
| L | M | N | O | H | I | J | K | D | E | F | G | | A | B | C |
| M | L | O | N | I | H | K | J | E | D | G | F | A | | C | B |
| N | O | L | M | J | K | H | I | F | G | D | E | B | C | | A |
| O | N | M | L | K | J | I | H | G | F | E | D | C | B | A | |

La utilización de la tabla anterior, representa las relaciones posicionales de los factores, por ejemplo: Si las columnas (1 y 2) son utilizadas para los factores A y B en estudio y es de interés en el estudio la interacción AB, ésta la colocaremos en la columna (3) ya que el cruce renglón A con columna B, se encuentra la letra C.

Otro ejemplo: Si en las columnas (8 y 10) tenemos ubicados factores los cuales nos interesa estudiar su interacción, entonces reservaremos la columna (2) para ubicar la interacción.

Otro ejemplo: Si en las columnas (5 y 10) tenemos ubicados factores los cuales nos interesa estudiar su interacción, entonces reservaremos la columna (15) para su interacción correspondiente.

5.2.1 Gráficas Lineales

Para la identificación de las posiciones de los factores y algunas interacciones en estudio de la tabla anterior, comúnmente utilizamos gráficas en línea recta, en las cuales cada **NODO** corresponde a un **FACTOR** y cada **LÍNEA** entre dos nodos corresponde a la **INTERACCIÓN** entre esos dos factores.

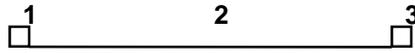
Así para el arreglo L_4 , tenemos:



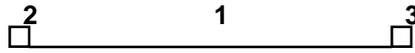
Columna 1: Colocar Factor A

Columna 2: Colocar Factor B

Columna 3: Colocar Interacción AB



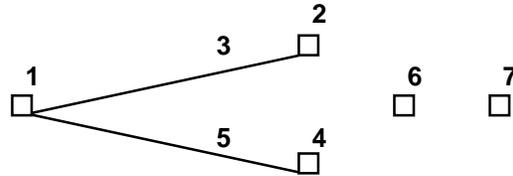
Columna 1: Colocar Factor A
 Columna 2: Colocar Interacción AB
 Columna 3: Colocar Factor B



Columna 1: Colocar Interacción AB
 Columna 2: Colocar Factor A
 Columna 3: Colocar Factor B

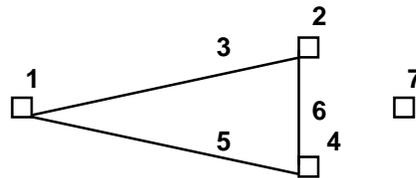
NOTA: Las 3 son válidas, analizar la tabla anterior.

Si deseamos estudiar 5 factores y dos interacciones, A, B, C, D, E, AB y AC, entonces la gráfica lineal puede corresponder a la siguiente:



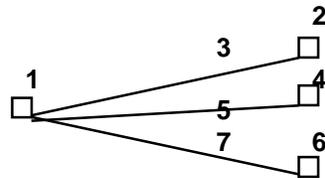
Columna 1: Factor A
 Columna 2: Factor B
 Columna 3: Interacción AB
 Columna 4: Factor C
 Columna 5: Interacción AC
 Columna 6: Factor D
 Columna 7: Factor E

Si deseamos estudiar 4 factores y tres interacciones, A, B, C, D, AB, AC y BC, entonces la gráfica lineal puede corresponder a la siguiente:



Columna 1: Factor A
 Columna 2: Factor B
 Columna 3: Interacción AB
 Columna 4: Factor C
 Columna 5: Interacción AC
 Columna 6: Interacción BC
 Columna 7: Factor D

O bien, A, B, C, D, AB, AC y AD



Columna 1: Factor A
 Columna 2: Factor B
 Columna 3: Interacción AB
 Columna 4: Factor C
 Columna 5: Interacción AC
 Columna 6: Factor D
 Columna 7: Interacción AD

5.2.2 Análisis Regular

Una forma para determinar que factores e interacciones pueden estar influyendo fuertemente a la variación y proyectar las condiciones óptimas, es la tabla de respuestas y la gráficas factoriales, las cuales se basan en las diferencias en el efecto promedio de los factores.

Este es el análisis de datos más básico para experimentos con arreglos ortogonales y se denomina análisis regular. Tal análisis involucra lo siguiente:

- 1.- Determinar la respuesta promedio de los niveles de los factores.
- 2.- Seleccionar los niveles óptimos de los factores comparando los promedios de las respuestas.
- 3.- Predecir el promedio del proceso para los niveles óptimos.
- 4.- Comparar los resultados de la predicción con lo obtenido en la corrida confirmatoria.

El análisis regular puede efectuarse a:

- 1.- Datos por variables
- 2.- Datos por atributos clasificados.

Análisis Regular para Datos por Variables

Una vez que se tengan los resultados de cada una de las pruebas, se procede de la siguiente forma:

- 1.- Determinar **N** (número total de datos), **T** (la suma total de los datos) y **X...** (el promedio de todos los datos).
- 2.- Obtener la suma de los datos y calcular el promedio por factor y por nivel.
- 3.- Obtener la suma de los datos y calcular el promedio por interacción y por nivel de interacción.
- 4.- Estimar el efecto que sea principal de cada uno de factores e interacciones, mediante la diferencia entre niveles.
- 5.- Elaborar una tabla de respuestas por factor e interacción, determinando el % de diferencia entre los niveles de cada factor e interacción por medio de la siguiente relación:
%DIFERENCIA = [(Resp. Mayor -- Resp. Menor)/Resp. Menor](100)
- 6.- Determinar cuáles son los factores e interacciones fuertes, las diferencias en los cambios de los efectos regirá a los factores e interacciones fuertes.
- 7.- Graficar los efectos de los factores e interacciones que resultaron fuertes a considerar. A medida que la pendiente de la línea sea mayor, más fuerte es el efecto.
- 8.- Al examinar las interacciones, se grafican los efectos promedios de cada una de las combinaciones posibles (A_1B_1 , A_1B_2 , A_2B_1 y A_2B_2 para el caso 2^2).
- 9.- Determinar el conjunto de niveles de los factores, los cuales se logra mejor respuesta, predecir la respuesta que se obtendrá con las recomendadas, la estimación de esta respuesta se logra al analizar los porcentajes de mejoramiento.
- 10.- Desarrollar corrida confirmatoria con el objeto de comprobar que los resultados son reproducibles; la corrida se lleva a cabo en las condiciones recomendadas.

En esta corrida confirmatoria, obtenemos R (resultado de la corrida confirmatoria), comparada con el promedio estimado a esas condiciones, podemos tener los siguientes resultados de R.

- a.-) **Si R es mucho mejor que el promedio estimado:** Significa que alguna interacción no incluida puede estar trabajando a nuestro favor.
- b.-) **Si R es poco mayor o menor que el promedio estimado:** Significa que hay buena probabilidad de reproducir los resultados.
- c.-) **Si R está alejada del promedio en el sentido desfavorable:** Significa que los datos experimentales no pueden ser aceptados, el experimento debe reconsiderarse.

Esta última comparación de R (la no deseada) puede deberse a las siguientes situaciones:

- 1.-) **Mala Aditividad:** Es equivalente a la existencia de interacciones fuertes y no hallan sido consideradas, también es probable que se hallan seleccionado factores de control con interacciones fuertes desfavorables.
- 2.-) **Insuficiencia de Factores de Control:** Esto puede provocar la no reproducibilidad de resultados en las condiciones seleccionadas. Tal vez no se tomó en cuenta algún factor fuerte.
- 3.-) **No identificación de Cambios:** Los niveles pueden haber sido muy parecidos para detectar cambios entre ellos.
- 4.-) **Presencia de Factores de Ruido:** La existencia de variaciones que inciden sobre el experimento y hacen que la respuesta sea inconsistente y poco fiable, por lo tanto respuestas con poca precisión.

Si por análisis de la corrida confirmatoria, descartamos Factores de Ruido, Insuficiencia de Factores de Control, entonces se deberá de buscar otras características de buena aditividad, en lugar de las interacciones presentes.

Análisis Regular para Atributos clasificados

Para este tipo de análisis, se requieren más repeticiones ya que son menos sensibles los atributos clasificados que las variables continuas. Para esta opción se procede en la siguiente forma:

- 1.- Determinar la respuesta para cada factor e interacción por nivel para cada atributo clasificado.
- 2.- Determinar el efecto de cada factor e interacción para cada atributo clasificado.
- 3.- Determinar la contribución de cada factor e interacción para atributo clasificado.
- 4.- Determinar en base al porcentaje de contribución, cuales son los factores con efectos más fuertes y las interacciones significativas.
- 5.- Para los factores más fuertes, determinar los porcentajes relativos acumulados de contribución por nivel de cada atributo clasificado.
- 6.- Elaborar la gráfica de porcentajes relativos acumulados de contribución por nivel para cada atributo clasificado.
- 7.- Determinar, en base a la gráfica y a la situación a optimizar, las condiciones que sugiere el experimento para ser recomendadas.
- 8.- Estimar la respuesta esperada, utilizando para ello únicamente las condiciones recomendadas (sugerencia de los niveles óptimos).

5.2.3 Relación Señal a Ruido

Para las tres opciones características de la función de pérdida de Calidad de Taguchi, se han generado expresiones de la relación Señal a Ruido; no hay que perder de vista que las relaciones Señal a Ruido consideran la tendencia central y la dispersión de los datos en una sola expresión y por lo tanto debemos de tener cuidado en el análisis de la información.

Tiene la ventaja de SER una expresión, la cual busca la maximización de dicha relación, pero las unidades las cuales manejamos los datos pueden influenciar los resultados.

No obstante, **ES MUY ÚTIL**, ya que si manejamos congruentemente las unidades de la información, nos permite análisis los cuales aumentamos el valor de la relación.

Las expresiones de la Relación Señal a Ruido manteniendo la siguiente convención: (NM => Nominal es Mejor), (mM => menor es Mejor) y (MM => Mayor es Mejor), son las siguientes:

$$\text{NM) } \left(\frac{S}{R_{\text{NM}}}_{\text{Mtra.}} \right) = 10 \text{Log} \left(\frac{\bar{X}}{S} \right)^2 = -20 \text{Log}(\text{CV})$$

$$\text{CV} = \frac{S}{\bar{X}} \text{ (Coeficiente de Variación)}$$

$$\text{mM) } \left(\frac{S}{R_{\text{mM}}}_{\text{Mtra.}} \right) = -10 \text{Log} \left(\frac{1}{n} \sum X_i^2 \right) = -10 \text{Log} \left[\bar{X}^2 + \frac{(n-1)}{n} S^2 \right]$$

$$\text{MM) } \left(\frac{S}{R_{\text{MM}}}_{\text{Mtra.}} \right) = -10 \text{Log} \left(\frac{1}{n} \sum \frac{1}{X^2} \right) = -10 \text{Log} \left[\bar{Y}^2 + \frac{(n-1)}{n} S_y^2 \right]$$

5.3 Caso Práctico de Aplicación

Para reafirmar la operatividad del Método de Taguchi, se presentará un caso desarrollado en una Empresa Galletera de la Región.

Se desea mejorar el grado de Aceptación del Público a un tipo de galleta que fabrica la Empresa, en la encuesta preliminar, se manejaron 5 conceptos que son: “**SABOR**”, “**GUSTO**”, “**CRUJIENTE**”, “**RESISTENCIA AL QUEBRADO**” y “**CONSISTENCIA**”.

En la encuesta reportaron que el “**SABOR**” y “**GUSTO**” deben mantenerse igual al existente, pero en los otros tres conceptos los ponderaron en la siguiente forma:

- 1) “**CRUJIENTE**” : Aproximadamente el 45% de personas encuestadas en 3 Ciudades.
- 2) “**RESISTENCIA AL QUEBRADO**” : Aprox. el 34% de personas encuestadas.
- 3) “**CONSISTENCIA AL COMER**” : Aprox. el 21% de las personas encuestadas.

5.3.1 Estudio y definición del problema

En la Empresa se tiene una escala continua de calificación (**X**) de **0 a 20**, en la cual **10** representa el valor meta Estándar de formalización de receta de fabricación, esta calificación (**X**) se obtiene a partir de un grupo evaluador que califica los 5 conceptos anteriores.

Manteniendo sin cambios “**SABOR**” y “**GUSTO**”, abocaremos nuestro estudio a los otros tres conceptos. Ya que en análisis técnicos podemos mantener sabor y gusto sin cambio aparente por medio de las grasas vegetales parcialmente hidrogenadas (Soya, Algodón, Palma y Cártamo); Lecitina; Enzimas; Jarabes y Azúcares.

La escala que maneja la Empresa, tiene puntos cualitativos de referencia únicamente, estos son:

- X = 0 : Galleta totalmente cruda en masa húmeda.
- X = 5 : Galleta cruda antes de empezar doramiento.
- X = 10 : Galleta conforme a receta estandarizada, => valor meta.
- X = 15 : Galleta excedida en doramiento.
- X = 20 : Galleta totalmente quemada.

El grupo evaluador, considera sus resultados calificando de **0 a 1** (**variables x's de cada concepto**), los tres conceptos analizados y la calificación final (**X**) se logra ponderando con los resultados de la encuesta, en la siguiente forma:

$$X = 20 [W_1X_1 + W_2X_2 + W_3X_3]$$

- En donde:
- X** => calificación final del lote de fabricación
 - W₁** = **.45** para este tipo de galleta.
 - W₂** = **.34** “ “ “ “ “
 - W₃** = **.21** “ “ “ “ “
 - X₁** => Concepto Crujiente.
 - X₂** => Concepto Resistencia al quebrado
 - X₃** => Concepto Consistencia.

La siguiente tabla presenta la escala apreciativa del grupo evaluador:

| | | X_1 | X_2 | X_3 |
|---------------|-----|------------------|---------------------------|----------------------------------|
| X_i | X | “Crujiente” | “Resistencia al Quebrado” | “Consistencia” plástica/elástica |
| de .00 a .20 | 0 | nulo (2) | nulo (2) | plástica (2) |
| de .21 a .40 | 5 | nulo (1) | nulo (1) | plástica (1) |
| de .41 a .59 | 10 | correcto | correcto | correcto |
| de .60 a .80 | 15 | excedido | excedido (1) | bajo |
| de .81 a 1.00 | 20 | exc. y carbonoso | excedido (2) | nulo |

La presentación de las galletas al público no se encuentra en estudio y continuará siendo en forma de paquete, se considera aceptable una variación de la calificación de $\pm 18\%$ con respecto al valor meta de $X = 10$ que es la media meta buscada.

Obteniendo datos de costeo, desglosando costos relacionados a la Calidad, y creando una correlación entre (X) y costos, se obtiene un costo en las tolerancias del lote de:

A = 82.70 \$/lote (costo equivalente a reprocesar, pero aquí es de no-aceptación)
Límite de Tolerancia Inferior => $X_{LTI} = 8.2$
Límite de Tolerancia Superior => $X_{LTS} = 11.8$
Valor Objetivo Meta => $m = 10$

NOTA: Los cálculos anteriores están basados en lotes de tamaño 450 Paq./Lote; cada paquete contiene 12 sobres, cada sobre contiene 2 galletas y aproximadamente cada galleta pesa 26 gramos. La Empresa tiene capacidad de manejar Lotes de fabricación hasta de 900 Paq./Lote, que corresponden entre 280 a 562 Kgr/Lote de producto terminado.

La relación lineal de cambio de **A**, con respecto al tamaño del Lote tuvo un coeficiente de determinación de $R^2 = .903$. Basado en 5 tamaños de lotes, $b_0 = 22.50$ y $b_1 = .133778$. Para lotes de 900 la **A** corresponde a ser de **142.9 \$/Lote**.

5.3.2 Determinación y definición del Objetivo

OBJETIVO: Se desea mejorar el grado de aceptación conforme a los resultados de la encuesta inicial, mediante una reformulación de la receta estandarizada.

5.3.3 Lluvia de ideas

En el análisis del proceso, y mediante las reuniones con las personas responsables se llegó a plantear los posibles Factores a considerar, algunos de ellos son fácilmente controlables en el proceso, otros no lo son tanto, pero para el estudio perfectamente pueden manejarse niveles de comparación, estos son:

FACT. INT. (CONTROLABLES)

- A)** Molde (vidrio, aluminio)
- B)** Velocidad de Batido (60, 90 RPM)
- C)** Tiempo de cocción (20, 40 min.)
- D)** Temperatura del Horno (170, 210 °C)

FACT. EXT. (NO-CONTROLABLES=> RUIDO)

- E)** Harina de Trigo Sarraceno Oscura
(bajo contenido de grasa => 25 mgr/Kgr,
alto contenido de grasa => 32 mgr/Kgr.)
- F)** Tiempo Enfriado antes de empaque.
(corto=> 4 min., largo=> 10 min.)
- G)** Humedad Relativa en el Enfriado (40, 80%)

5.3.4 Diseño del Experimento

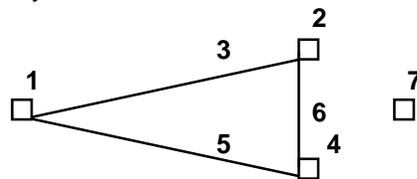
Mediante la tabla de interacciones de la sección 5.2 y considerando de interés el estudio de los efectos **A, B, C, D, BC, BD, CD** por el análisis regular aplicado al proceso, se decide un arreglo $L_8(2^7)$ para los factores internos.

Para los factores externos no se consideran de interés el estudio de las interacciones, por lo que se decide un arreglo $L_4(2^3)$, colocando los efectos en el siguiente orden **E, F, G**.

Para el arreglo L_8 , se tiene el siguiente acomodo de efectos cumpliendo la tabla estándar de interacciones.

| COLUMNAS | | | | | | |
|----------|---|----|---|----|----|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| B | C | BC | D | BD | CD | A |

La gráfica lineal que corresponderá a este Diseño de Experimento se está estudiando 4 factores y 3 interacciones de interés.



- Columna 1: Factor B
- Columna 2: Factor C
- Columna 3: Interacción BC
- Columna 4: Factor D
- Columna 5: Interacción BD
- Columna 6: Interacción CD
- Columna 7: Factor A

5.3.5 Generación y conducción del Experimento

Como se tienen 8 corridas en los factores internos y 4 en factores externos, se desarrollarán $8 \times 4 = 32$ corridas a las condiciones de los arreglos tanto en los ensayos preliminares como en los ensayos confirmatorios, se obtuvieron los siguientes datos preliminares con el grupo evaluador:

| | | | | | | | | G | 1 | 2 | 2 | 1 | | | |
|-------|---|---|----|---|----|----|---|------|------|------|------|------|--|--|--|
| | | | | | | | | F | 1 | 2 | 1 | 2 | | | |
| | | | | | | | | E | 1 | 1 | 2 | 2 | | | |
| L_8 | B | C | BC | D | BD | CD | A | 1 | 2 | 3 | 4 | Tot | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8.1 | 7.9 | 10.3 | 10.1 | 36.4 | | | |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 10.4 | 10.2 | 8.4 | 8.0 | 37.0 | | | |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 12.3 | 12.0 | 9.2 | 10.3 | 43.8 | | | |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 13.7 | 13.5 | 12.9 | 12.0 | 52.1 | | | |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 8.3 | 8.0 | 8.0 | 7.6 | 31.9 | | | |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 8.6 | 8.3 | 8.1 | 8.2 | 33.2 | | | |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 9.3 | 8.8 | 8.0 | 8.6 | 34.7 | | | |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 10.3 | 9.9 | 9.3 | 9.7 | 39.2 | | | |

Los niveles 1 y 2 corresponden a las dos situaciones que en cada Factor es necesario considerar.

5.3.6 Análisis y conclusiones de resultados preliminares

Los totales, promedios, SS, GL, CM, Fo y porcentaje son:

| Col. | Factor | Tot 1 | Tot 2 | Med 1 | Med 2 | SS | GL | CM | Fo | (%) |
|---------------|----------|--------|--------|-----------------|---------|---------|----|---------------|---------|----------|
| 1 | B-Vbatid | 169.3 | 139.0 | 10.5813 | 8.6875 | 28.6903 | 1 | 28.6903 | 35.6355 | 29.08 % |
| 2 | C-Tcocci | 138.5 | 169.8 | 8.6563 | 10.6125 | 30.6153 | 1 | 30.6153 | 38.0265 | 31.09 % |
| 3 | BCInter | 147.3 | 161.0 | 9.2063 | 10.0625 | 5.8653 | 1 | 5.8653 | 7.2852 | 5.28 % |
| 4 | DThorno | 146.8 | 161.5 | 9.1750 | 10.0938 | 6.7528 | 1 | 6.7528 | 8.3875 | 6.20 % |
| 5 | BDInter | 152.6 | 155.7 | 9.5375 | 9.7313 | 0.3003 | 1 | 0.3003 | 0.3730 | - 0.53 % |
| 6 | CDInter | 159.6 | 148.7 | 9.9750 | 9.2938 | 3.7128 | 1 | 3.7128 | 4.6116 | 3.03 % |
| 7 | A-molde | 156.4 | 151.9 | 9.7750 | 9.4938 | 0.6328 | 1 | 0.6328 | 0.7860 | - 0.18 % |
| Resid. | XXXXX | XXXXXX | XXXXXX | XXXXXX | XXXXXX | 19.3225 | 24 | 0.8051 | XXXXXX | 26.03 % |
| Total | XXXXX | XXXXXX | XXXXXX | 9.634375 | | 95.8922 | 31 | XXXXXX | XXXXXX | 100 % |

A un nivel de significación del 5%, $F_c = 4.26$ las significaciones son (1-SI), (2-SI), (3-SI), (4-SI), (5-NO), (6-SI) y (7-NO).

Los efectos B, C, BC, D y CD han sido influyentes en el cambio de la calificación, las interacciones se pueden controlar con modificación de los niveles de los factores, encontrando una adecuada gráfica que veremos con el programa de Taguchi que se incluirá.

Por medio de un análisis de Señal a Ruido y su análisis de variaciones de la misma obtenemos:

| Col | Efecto | S/R _{m1} | S/R _{m2} | SS _{S/R} | GL | CM _{S/R} | (%) |
|--------------|----------|--------------------|-------------------|-------------------|----|-------------------|----------|
| 1 | B-Vbatid | 19.1884 | 28.0378 | 156.623 | 1 | 156.623 | 69.33 % |
| 2 | C-Tcocci | 23.7954 | 23.4308 | 0.26593 | 1 | 0.26593 | 0.12 % |
| 3 | BCInter | 21.5524 | 25.6738 | 33.9708 | 1 | 33.9708 | 15.04 % |
| 4 | DThorno | 21.8957 | 25.3305 | 23.5962 | 1 | 23.5962 | 10.44 % |
| 5 | BDInter | 23.4215 | 23.8047 | 0.29373 | 1 | 0.29373 | 0.13 % |
| 6 | CDInter | 24.5041 | 22.7221 | 6.35036 | 1 | 6.35036 | 2.81 % |
| 7 | A-molde | 24.389 | 22.8372 | 4.81565 | 1 | 4.81565 | 2.13 % |
| Total | xxxxxxxx | 23.61310129 | | 225.916 | 7 | xxxxxxxx | xxxxxxxx |

Mediante la discusión de resultados y observando que en Señal a Ruido se es más sensible en velocidad de batido, se reducirán los niveles de 60, 90 a 70, 80 RPM; en tiempo de cocción de 20, 40 a 25, 35 minutos; en temperatura del horno de 170, 210 a 180, 200 grados centígrados, puesto que el Molde no fue significativo.

5.3.7 Corridas confirmatorias de conclusiones

A las nuevas condiciones obtenidas de las conclusiones de la preliminar, se generó nuevamente el muestreo arrojando la siguiente información:

| | | | | | | | | G | 1 | 2 | 2 | 1 | | |
|----------------|---|---|----|---|----|----|---|------|------|------|------|------|--|--|
| | | | | | | | | F | 1 | 2 | 1 | 2 | | |
| | | | | | | | | E | 1 | 1 | 2 | 2 | | |
| L ₈ | B | C | BC | D | BD | CD | A | 1 | 2 | 3 | 4 | Tot | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 9.9 | 9.8 | 10.1 | 10.0 | 39.8 | | |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 10.2 | 10.1 | 9.8 | 9.9 | 40.0 | | |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 10.2 | 10.1 | 10.1 | 9.9 | 40.3 | | |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 10.1 | 10.3 | 10.2 | 10.0 | 40.6 | | |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 9.8 | 9.8 | 10.0 | 9.9 | 39.5 | | |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 10.1 | 10.1 | 9.9 | 10.0 | 40.1 | | |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 9.8 | 9.8 | 10.0 | 10.1 | 39.7 | | |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 10.0 | 9.9 | 9.9 | 10.0 | 39.8 | | |

Los totales, promedios, SS, GL, CM, Fo y porcentaje son:

| Col. | Factor | Tot 1 | Tot 2 | Med 1 | Med 2 | SS | GL | CM | Fo | (%) |
|--------|----------|--------|--------|----------------|---------|--------|----|---------------|--------|----------|
| 1 | B-Vbatid | 160.7 | 159.1 | 10.0438 | 9.9438 | 0.0800 | 1 | 0.0800 | 5.0526 | 10.72 % |
| 2 | C-Tcocci | 159.4 | 160.4 | 9.9625 | 10.0250 | 0.0312 | 1 | 0.0312 | 1.9737 | 2.57 % |
| 3 | BCInter | 159.3 | 160.5 | 9.9563 | 10.0313 | 0.0450 | 1 | 0.0450 | 2.8421 | 4.87 % |
| 4 | DThorno | 159.3 | 160.5 | 9.9563 | 10.0313 | 0.0450 | 1 | 0.0450 | 2.8421 | 4.87 % |
| 5 | BDInter | 160.0 | 159.8 | 10.0000 | 9.9875 | 0.0013 | 1 | 0.0013 | 0.0789 | - 2.44 % |
| 6 | CDInter | 159.7 | 160.1 | 9.9813 | 10.0063 | 0.0050 | 1 | 0.0050 | 0.3158 | - 1.81 % |
| 7 | A-molde | 160.2 | 159.6 | 10.0125 | 9.9750 | 0.0112 | 1 | 0.0112 | 0.7105 | - 0.77 % |
| Resid. | XXXXX | XXXXXX | XXXXXX | XXXXXX | XXXXXX | 0.3800 | 24 | 0.0158 | XXXXXX | 81.98 % |
| Total | XXXXX | XXXXXX | XXXXXX | 9.99375 | | 0.5988 | 31 | XXXXXX | XXXXXX | 100 % |

| Col. | Factor | S/N _{m1} | S/N _{m2} | SS _{S/R} | GL | CM _{S/R} | (%) |
|-------|----------|--------------------|-------------------|-------------------|----|-------------------|----------|
| 1 | B-Vbatid | 37.1223 | 40.4522 | 22.176 | 1 | 22.176 | 34.52 % |
| 2 | C-Tcocci | 38.2943 | 39.2801 | 1.94349 | 1 | 1.94349 | 3.03 % |
| 3 | BCInter | 38.4124 | 39.1621 | 1.12392 | 1 | 1.12392 | 1.75 % |
| 4 | DThorno | 38.1221 | 39.4523 | 3.53877 | 1 | 3.53877 | 5.51 % |
| 5 | BDInter | 40.2336 | 37.3409 | 16.7358 | 1 | 16.7358 | 26.05 % |
| 6 | CDInter | 40.1613 | 37.4132 | 15.104 | 1 | 15.104 | 23.51 % |
| 7 | A-molde | 38.1153 | 39.4592 | 3.61204 | 1 | 3.61204 | 5.62 % |
| Total | xxxxxxxx | 38.78723222 | | 64.2341 | 7 | xxxxxxxx | xxxxxxxx |

Obviamente en el ANDEVA original de la corrida confirmatoria (esta última), ahora únicamente esta siendo significativo el efecto B que es velocidad de Batido, pero la Empresa prefiere mantener actualmente los valores 70,80 RPM ya que un control más preciso ocasionará costos adicionales no aceptables.

5.3.8 Determinación del valor esperado de Costo

El costo por lote lo obtenemos con las ecuaciones al principio de Unidad, teniendo en cuenta que el gran promedio estima el valor de la media μ . Estimado lector al obtener los totales y sacar su media es de: **9.993751**. El valor de **A** es **82.70 \$/lote** para lotes de 450 paquetes y **142.90 \$/lote** para lotes de 900 paquetes.

Aplicando las siguientes fórmulas obtenemos los valores:

$$\delta = (LSE - m) = (m - LIE)$$

$$k = A/\delta^2$$

$$E(L(x)) = k(\sigma^2 + D^2)$$

$$D^2 = (\mu - m)^2$$

| Lotes | A | δ | k | D^2 | σ^2 | Costo/lote |
|----------|--------|----------|---------|------------|------------|------------|
| 450 Paq. | 82.70 | 1.8 | 25.5247 | 0.00003905 | 0.0158 | 0.4043 |
| 900 Paq. | 142.90 | 1.8 | 44.1049 | 0.00003905 | 0.0158 | 0.6986 |

Estimado lector, todos los datos están dentro de las especificaciones para las galletas, se están cumpliendo perfectamente los límites y además el Costo por Lote se redujo, por favor fuera del curso estime el valor del costo antes del método de Taguchi.

5.3.9 Conclusiones Finales

Se está teniendo menos variabilidad (más precisión, de **0.8051 a 0.0158**) en los nuevos niveles para los efectos y las interacciones desfavorecedoras disminuyen, se ajusta mejor al valor meta de 10 (de **9.634376 a 9.993751**), y a su vez el costo por Lote disminuye (**se le pidió que lo estime**), **CONCLUSIÓN FINAL ==> "LA RECETA OPERATIVA DE FABRICACIÓN DE LA GALLETA ES MÁS ROBUSTA Y MENOS COSTOSA"**.

5.4 Bibliografía sugerida



1) DISEÑO Y ANÁLISIS DE EXPERIMENTOS: para todas las áreas, el autor es: **Douglas C. Montgomery**, Arizona State University (original en Inglés 1991); traducción por el Lic. Jaime Delgado Saldivar (1993), Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) Campus León Gto., México; revisión técnica por la Lic. Bertha Dávila de Apodaca, Depto. de Matemáticas (ITESM) Campus Monterrey N.L., México; **Grupo Editorial Iberoamérica.**



2) FUNDAMENTAL CONCEPTS IN THE DESIGN OF EXPERIMENTS: para todas las áreas, el autor es: **Charles R. Hicks**, Purdue University (Second Edition); Lafayette, Indiana, March 1973, editorial; **Holt, Rinehart and Winston.**



3) PROBABILITY AND STATISTICS In Engineering and Management Science: para las áreas de Ingenierías, áreas Técnicas y Económico Administrativas, los autores son: **William W. Hines & Douglas C. Montgomery**, Georgia Institute of Technology ambos autores (1973), editorial **Ronald Press Company, New York.**



4) EXPERIMENTAL DESIGN a Chemometric approach: para las áreas de Ingeniería Química y Química, los autores son: **Stanley N. Deming** (Department of Chemistry, University of Houston, Houston, TX, U.S.A.) y **Stephen L. Morgan** (Department of Chemistry, University of South Carolina, Columbia, SC, U.S.A.) (1988), distribuido para U.S.A. y Canada por **Elsevier Science Publishing Company, 655, Avenue of the Americas, New York, NY 10010 U.S.A., Printed (1987) in Amsterdam, The Netherlands.**



5) QUALITY ENGINEERING IN PRODUCTION SYSTEMS: para todas las áreas, los autores son: **Genichi Taguchi** (International Consultant), **Elsayed A. Elsayed** (Department of Industrial Engineering, Rutgers University), **Thomas C. Hsiang** (Director of Statistical Services, Universal Foods Corporation) (1989), editorial **McGraw-Hill Company.**



6) TAGUCHI ON ROBUST TECHNOLOGY DEVELOPMENT Bringing Quality Engineering Upstream: para todas las áreas, el autor es: **Genichi Taguchi (1990)**, originalmente publicado por **Central Japan Quality Control**, traducido al Inglés por Shih Chung Tsai (1993) publicado por **ASME PRESS, New York.**



7) ESTRATEGIAS EXPERIMENTALES PARA EL MEJORAMIENTO DE LA CALIDAD EN LA INDUSTRIA: para las Ingenierías, sus autores son: **John Lawson / José L. Madrigal / John Erjavec**, (1992), **Grupo Editorial Iberoamérica.**



8) ADMINISTRACIÓN Y CONTROL DE LA CALIDAD: para todas las áreas, los autores son: **James R. Evans / William M. Lindsay**, (original 1993, University of Cincinnati & Northern Kentucky University, publicado por West Publishing Company), traducido, editado y distribuido (Enero 1995) por **Grupo Editorial Iberoamérica.**