CAPITULO VI

DISEÑO DE EXPERIMENTOS

6.1. Generalidades

En 1964 las compañías americanas generaban alrededor de 6 mil millones de dólares en superávit comercial de exportación, sin darse cuenta que estaban bajo el ataque de compañías extranjeras. Ya por el año de 1984 en los Estados Unidos se producía un déficit de la balanza comercial de 123 mil millones de dólares. En este período de 20 años la productividad en los Estados Unidos se incremento tan solo en un 35 %, comparado con un 60 % de algunos países europeos, y con un sorprendente *120 % en el Japón.* El mismo país que tenía la reputación de producir "chatarra" en los años 50 - 60, ahora amenazaba la supervivencia de la industria americana al producir productos de *alta calidad y bajos precios.*

Como ocurrió aquello? Es bien conocido que mucho de este sorprendente cambio en la calidad de los productos japoneses se debió a controles de calidad de los productos introducidos mediante el uso de el *diseño de experimentos* y el control de calidad durante la producción mediante técnicas estadísticas. Irónicamente, estas técnicas fueron desarrolladas mayormente en los Estados Unidos hace unos 60 años atrás, pero las compañías americanas no se preocuparon en implementar estas técnicas. Esta despreocupación era debida principalmente a la escasa competencia que existía después de la postguerra lo que no hacia útil el preocuparse excesivamente por controles de calidad. Después de la Segunda Guerra Mundial, los japoneses eran muy receptivos a las ideas de los estadísticos y como resultado, los japoneses llegaron a ser expertos en calidad y pronto sus empresas prosperaron. Así, todos en una compañía, desde los gerentes hasta los operarios de línea, fueron adiestrados en estas técnicas, no solo para controlar la calidad de producción , sino para mejorar continuamente los procesos productivos.

6.2. Qué es un diseño experimental?

Los investigadores e ingenieros realizan experimentos virtualmente en todos los campos del saber, por lo general para descubrir algo acerca de un proceso o sistema en particular. Literalmente un experimento es una *prueba o ensayo*. Un experimento diseñado es una prueba o serie de pruebas en las cuales se inducen *cambios deliberados en las variables de entrada de un proceso o sistema*, de manera que sea posible observar e identificar las causas de los cambios en el trabajo de *optimización.*

El proceso o sistema bajo estudio puede representarse por medio del modelo de la Figura 6.1. Suele ser posible visualizar el proceso como una combinación de una serie de recursos (máquinas, métodos, personas, etc.) que transforman alguna entrada (por lo general un material) en una salida que tiene una ó más respuestas observables. Algunas variables del proceso son controlables, x_1, x_2,..., x_p, mientras que otras son z_1, z_2, ..., z_p son incontrolables (aunque en un experimento se la puede controlar). Por ejemplo, en un proceso de endurecimiento por templado de un acero, el tiempo y temperatura son variables controlables; mientras que el contenido de carbono y elementos aleantes son variables incontrolables.

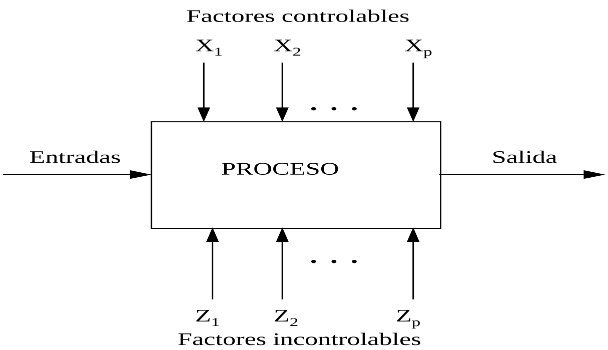

Figura 6.1: Modelo general de un proceso o sistema.

En esencia, un diseño experimental es un acercamiento científico que permite a un investigador entender mejor un proceso y a determinar la manera en que las variables de ingreso afectan una respuesta o variables de salida. Las condiciones de cada una de las variables de ingreso tienen efecto en las respuestas de producción; algunas de ellas tendrán un efecto más pronunciado que otras. Además, estas variables interactuan entre si, es decir, se desarrolla entre ellas una interdependencia que modifica su efecto individual. Un aporte tan importante como el de evaluar el efecto de cada variable, es que un diseño experimental también permite medir el nivel de interacción entre las variables de ingreso y fijar los mejores valores de estas variables de modo que la respuesta proporcione las mejores condiciones.

Supóngase por ejemplo, que un ingeniero esta interesado en estudiar el efecto que tienen sobre una aleación de aluminio dos procesos diferentes de endurecimiento: a) el templado en aceite y b) el templado en agua salada. En este caso el objetivo del investigador es determinar cual de las dos soluciones produce el máximo grado de dureza sobre la aleación mencionada. El ingeniero debe someter un cierto número de probetas de la aleación a cada medio de templado, después medir la dureza de las muestras. La dureza promedio de las probetas tratadas en cada solución servirá para determinar cuál de las dos soluciones es la mejor.

Al diseñar el experimento en mención, viene a la mente algunas preguntas importantes;

a) Son estas soluciones los únicos medios de templado?

b) Existen otros factores que puedan afectar la dureza de las muestras? Tipo de probeta por ejemplo.

c) Cuántas probetas de debe considerar para el estudio?

d) Que diferencia en los niveles promedio de dureza entre las dos soluciones debe considerarse importante?

En cualquier experimento, los resultados y conclusiones que pueden obtenerse dependen, en gran parte, de la forma en que los datos fueron recopilados. Por ejemplo, supóngase que en las pruebas de templado se usan probetas de un tipo para templado en aceite, y de otro tipo para templado en agua. Luego al comparar los resultados, no será posible decir cuanta de la diferencia observada se debe a la diferencia de los medios de templado y cuanta al tipo de probeta. De este modo, el método utilizado en la obtención de los datos ha afectado las conclusiones que pueden deducirse del experimento.

6.3. Tipos de experimentación.

Ahora bien, la manera como pueden ser llevados a cabo los experimentos pueden ser de dos formas:

- Experimentación pasiva, denominada también convencional o clásica.
- Experimentación activa, conocida como diseños experimentales.

6.3.1 Experimentación pasiva

En este tipo de experimentación los experimentos son llevados a cabo variando una variable a la vez. Esto es por ejemplo en el caso de dos variables controlables, en una primera corrida experimental se hace variar la primera variable mientras que la segunda se mantiene constante, y después variar la segunda variable mientras que la primera se mantiene constante.

Consideremos el siguiente ejemplo: un jefe de planta conoce que la temperatura y presión de producción de ciertos artículos tienen efecto en la cantidad de artículos deficientes. Decide entonces diseñar unos experimentos para probar la el efecto de temperatura y presión en la cantidad de productos defectuosos. El diseño consiste en (1) seleccionar dos niveles de temperatura, T_1 y T_2 y dos niveles de presión, P_1 y P_2; y, (2) Mantener la presión P_1 fija y variar la temperatura en sus dos niveles fijados. Hecha la experimentación se observan los registrados en la Tabla 6.1.

Tabla 6.1: Aproximación uno-a-la-vez (parte I)

Presión	Temperatura	Respuesta
P_1	T_1	0.030
P_1	T_2	0.015

Se concluye de esto que la combinación P_1 - T_2 proporciona menor cantidad de defectos. Ya que la temperatura T_2 es la que mejor responde, se prueba variaciones de presión a T_2. Los resultados se observan en la Tabla 6.2:

Tabla 6.2: Aproximación uno- a- la- vez (parte II)

Presión	Temperatura	Respuesta
P_1	T_2	0.015
P_2	T_2	0.010

A primera vista se concluye que la combinación P_2 - T_2 proporciona el más bajo índice de defectos, 1%. Este tipo de experimentación se denomina "*diseño uno a la vez*". Este método no es muy conveniente cuando se investiga dos o más variables, ya que no considera la posible *interacción* que ocurre entre las variables, teniendo además la probabilidad de conducirnos a un falso óptimo.

6.3.2 Experimentación activa

Los experimentos son llevados a cabo mediante un diseño predeterminado (diseño de experimentos), el cual es un plan organizado de experimentos que permiten evaluar *simultáneamente* todas las variables consideradas y además, evaluar la fuerza de interacción entre las variables y reducir el número de experimentos a llevarse a cabo.

En el ejemplo anterior, supóngase que combinación no probada P_2 - T_1 produce una interacción como se observa en la Figura 6.2. La interacción se interpreta del siguiente modo: cuando la temperatura cambia de *T1* a *T2* el cambio en los productos defectuosos es diferente, dependiendo de los valores de presión fijados. Es de observarse que el diseño apropiado debería probar todas las cuatro combinaciones de presión y temperatura, y evaluar las interacciones, tal como se hace en un diseño factorial completo.

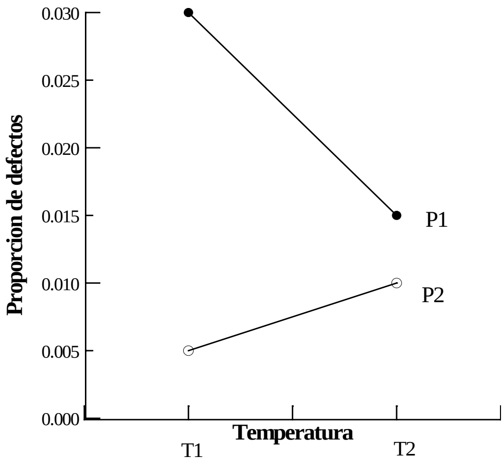

Figura 6.2. Gráfico de interacción Presión - Temperatura.

En la optimización de los procesos, la utilización de los diseños experimentales, junto con una estrategia adecuada de optimización, es el mejor método para encontrar los valores óptimos de las variables de la manera más rápida y eficiente posible.

6.4. Aplicaciones de los Diseños Experimentales

Los métodos de diseño experimental tienen amplia aplicación en muchas disciplinas. En efecto, es posible considerar a la experimentación parte del proceso de i*nvestigación científica* y una de las formas en que aprendemos acerca de la forma en que funcionan los sistemas o procesos.

El diseño experimental es un medio de importancia crítica en el medio de la ingeniería para mejorar el rendimiento de un proceso. También se emplea extensamente en el desarrollo de nuevos procesos. La aplicación de técnicas de diseño experimental en una fase temprana de un proceso:

- Da la mayor información por experimento, que la experimentación clásica. El investigador que hace uso de esta técnica reduce el tiempo de experimentación y mejora su eficiencia, especialmente cuando son muchas las variables de potencial importancia.
- Permite una colección y análisis organizado de la información obtenida, que conduce a conclusiones confiables mediante un análisis estadístico.
- Gran confiabilidad de los resultados.
- Permite evaluar las interacciones entre las variables experimentales, y que conducen a predicciones más confiables de las respuestas en áreas no directamente cubiertas por la experimentación.
- Mejora del rendimiento del proceso.
- Menor variabilidad y mayor apego a los requerimientos u objetivo.
- Menor tiempo de desarrollo.
- Menores costos globales.

El uso de diseño experimental en las tareas de desarrollo de nuevos procesos y diseño de ingeniería, puede dar por resultado productos con mayor confiabilidad y mejor funcionamiento, menores costos, y menor tiempo de diseño y desarrollo del producto.

6.5. Principios básicos

Para que un experimento se realice en la forma más eficiente, es necesario emplear métodos científicos en su ***planeación*.** El diseño estadístico de experimentos es el proceso de planear un experimento para obtener datos apropiados, que pueden ser analizados mediante métodos estadísticos, con el objeto de producir conclusiones *válidas y objetivas*. La metodología estadística es el único enfoque objetivo para analizar un problema que involucre datos sujetos a errores experimentales. De ese modo, es establecen dos aspectos en cualquier problema experimental:

- el diseño del experimento; y,
- el análisis estadístico de los datos.

Los tres principios básicos en el diseño de experimentos son el de:

a) Obtención de réplicas;
b) Aleatorización; y
c) Análisis por bloques.

La ***réplica*** se refiere a la repetición del experimento básico. Este concepto tiene dos propiedades importantes. En primer lugar permite obtener una estimación del error experimental; tal estimación se convierte en la unidad básica para determinar si las diferencias observadas en los datos son *estadísticamente* significativas. En segundo lugar, el uso de réplicas permite al experimentador calcular una estimación más precisa del efecto de un factor en el experimento, si se usa por ejemplo la media, $\bar{y}$, como una estimación de dicho efecto; esto porque la varianza, de un muestra de tamaño *n*, tendrá una varianza igual a σ^2/n , lo que implica que la variabilidad de una respuesta experimental tiende a cero cuando *n* es más grande.

La ***aleatorización*** es la piedra angular que fundamenta el uso de métodos estadísticos en el diseño de experimentos. Se entiende por aleatorización el hecho de que tanto la asignación del material experimental como el orden en que se realizan las pruebas individuales se determinan aleatoriamente. Los métodos estadísticos requieren que las observaciones (o los errores) sean variables aleatorias independientes. Además, al aleatorizar adecuadamente un experimento se ayuda a 'cancelar' los efectos de factores extraños que pudieran estar presentes.

El ***análisis por bloques*** es una técnica que se usa para incrementar la precisión del experimento. Un bloque es una porción del material experimental que sea más homogénea que el total del material. Al realizar un análisis por bloques. se hacen las comparaciones entre las condiciones de interés del experimento dentro de cada bloque.

6.6. Directrices para el diseño de experimentos

Para usar un enfoque estadístico al diseñar y analizar un experimento, se requiere que todos los participantes en él tengan de antemano una idea clara de qué es exactamente lo que se va a estudiar, cómo se van a recopilar los datos y , al menos, una idea cualitativa de cómo se van a analizar. A continuación se ofrece una guía del procedimiento recomendado.

A) Comprensión y planteamiento del problema

Aunque puede parecer obvio, en la práctica no es sencillo darse cuenta de que existe un problema que requiere experimentación, ni diseñar un planeamiento claro y aceptable. Es necesario desarrollar todas las ideas sobre los objetivos del experimento. Suele ser importante solicitar la opinión de todas las partes implicadas, quienes pueden aportar con conocimiento detallado de algún proceso. Un planteamiento claro del problema contribuye a un mejor conocimiento del fenómeno y de la solución final del problema.

B) Elección de factores y niveles

La elección de los factores o variables, intervalos y niveles específicos a los cuales se hará el experimento, son tareas que deben emprenderse desde el inicio. También la forma de controlar esos factores y los métodos de medición. Es importante fijar todos los factores que pueden ser de interés, y no depender demasiado de experiencias pasadas, en particular durante las primeras etapas, cuando el objetivo es la caracterización del proceso.

C) Selección de la variable de respuesta

Al seleccionar la respuesta o variable dependiente, se debe estar seguro que la respuesta que se va a medir realmente provea información útil acerca del proceso de estudio. Usualmente, el promedio o la desviación estándar (ó ambos) de la característica medida, serán la variable de respuesta. La capacidad de medición es también un factor importante, y si esta es deficiente, no podrá esperarse mas que la detección de efectos relativamente grandes.

D) Elección del diseño experimental

Para elegir el diseño que se va a emplear, es necesario considerar el tamaño de muestra (número de repeticiones), seleccionar el orden adecuado para los ensayos experimentales, y determinar si hay implicado bloqueo u otras restricciones de aleatorización.

E) Realización del experimento

Es vital aquí, verificar que el proceso de experimentación se desarrolle de acuerdo a lo planificado. En esta fase, los errores en el procedimiento suelen anular la validez experimental.

F) Análisis de datos

Se requiere de un análisis estadístico de los datos, los cuales por lo general no son complicados. Se debe tener en cuenta que el análisis estadístico solo hace posible obtener el probable error de una conclusión o asignarle un nivel de confiabilidad a los resultados. Las técnicas estadísticas, aunadas a un buen conocimiento técnico del proceso y al sentido común, suelen llevar a conclusiones razonables.

H) Conclusiones y recomendaciones

Para esta etapa, es muy útil utilizar métodos gráficos. También es importante realizar corridas de seguimiento y pruebas de confirmación para dar validez a las conclusiones del experimento.

6.7. Clasificación de los diseños experimentales

La gran variedad de aplicaciones de los diseños experimentales (caracterización de procesos, diseño de productos, optimización de procesos, aplicaciones en ingeniería de calidad) no ha permitido una clasificación bien definida de estos. Desde el punto de vista de la optimización se pueden clasificar teniendo en cuenta el orden del modelo matemático a obtenerse (diseños de primer o segundo orden). Desde el punto de vista de la manera en que son llevados los diseños experimentales pueden clasificarse en simultáneos y secuenciales. Desde el punto de vista del objetivo (caracterización de procesos, diseño de productos, etc.) se tiene diseños por comparación simple hasta diseños jerárquicos o anidados.

6.7.1 Experimentos por comparación simple

Se usan para comparar dos condiciones (a menudo tratamientos). El objetivo es analizar si dos fórmulas diferentes de un producto producen resultados equivalentes. Este tratamiento requiere la aplicación de conceptos estadísticos tales como variables aleatorias, distribución de probabilidades, distribuciones muestrales y pruebas de hipótesis.

6.7.2 Diseños experimentales de primer orden

Son experimentos que conducen a optimización. Permiten estimar adecuadamente los coeficientes β de la ecuación o modelo matemático empírico de la siguiente expresión:

$$\hat{Y} = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Donde $\hat{Y}$ es la respuesta o variable de salida (variable dependiente) que es función de las variables de entrada $X_1, X_2, \ldots, X_k$. El más simple caso, con una sola variable de entrada, k = 1, proporciona un modelo como el siguiente:

$$\hat{Y} = \beta_o + \beta_1 X_1$$

El modelo permite *predecir* los valores de *Y* sobre un *rango limitado de valores de X*.

Para dos o más variables, la situación se complica por la posible existencia de interacciones entre variables, lo que indica que las variables no actúan independientemente sobre la variable respuesta. el modelo empírico para dos variables será:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$$

donde β_{12} es el denominado parámetro de interacción. Cuando β_{12} es cero, se obtiene un modelo estrictamente de primer orden con dos líneas rectas paralelas entre si. Cuando β_{12} es diferente de cero, el modelo es interactivo y las líneas no son paralelas y se interceptan entre si.

Los diseños experimentales que permiten estimar modelos de primer orden son:

- Diseños factoriales;
- Diseños factoriales fraccionados;
- Diseños de *Placckett y Burman*; etc.

La utilización de diseños de primer orden son de gran importancia en la etapa inicial de optimización, donde al inicio de la investigación se tiene gran número de variables . El objetivo es estudiar el efecto de cada una de esas variables en la respuesta del proceso, eliminando aquellas que resulten poco significativas o con pendientes cercanas a cero.

6.7.3 Diseños experimentales de segundo orden

Se denomina así a aquellos que permiten estimar los parámetros β_j, β_{uj} , β_{jj} de la ecuación general:

$$\hat{Y} = \beta_0 + \sum_{j=1}^{k} \beta_j X_j + \sum_{uj=1}^{k} \beta_{uj} X_u X_j + \sum_{j=1}^{k} \beta_{jj} X_j^2$$

Si consideramos el caso de una sola variable controlable, el modelo de segundo orden toma la forma:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_{11} X^2$$

donde el término β_{11} es denominado el término de curvatura y el punto máximo corresponde a $-\beta_1/2\beta_{11}$

Si consideramos ahora el caso de dos variables, el modelo de segundo orden interactivo toma la siguiente forma:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$$

donde β_{11} y β_{22} son los parámetros de curvatura y β_{12} es el parámetro de interacción; la Figura 6.3 muestra en forma gráfica una ecuación como esta, en la cual se puede observar la región óptima.

Los diseños experimentales más comunes que permiten estimar modelos de segundo orden son:

- Diseños rotables: hexagonal, octogonal, etc.;

- Diseños compuestos: compuesto central, compuestos de más de tres variables, etc.

Estos diseños de segundo orden son indispensables en la etapa final de optimización, donde un modelo de estos puede descubrir aproximadamente la región óptima.

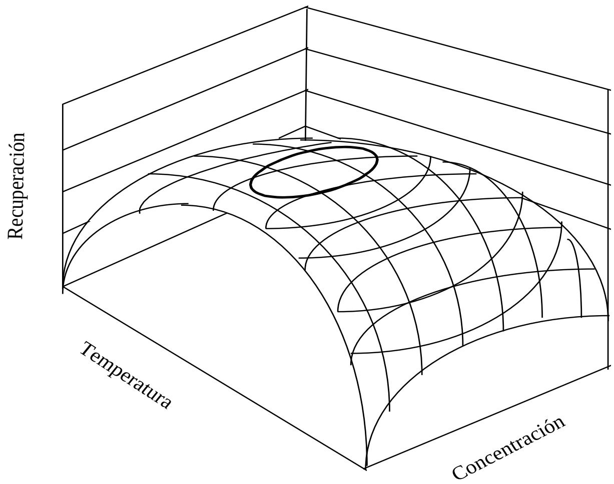

Figura 6.3: Representación gráfica de una ecuación de segundo orden, tipo:

$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$ producto de un diseño hexagonal.

6.8. Modelos matemáticos

Un modelo matemático es un sistema de ecuaciones algebraicas o diferenciales que representan cuantitativamente el proceso o algunos aspectos del proceso. Su formulación depende del conocimiento de las leyes que gobiernan el proceso y de la habilidad de expresarlas matemáticamente.

Por ejemplo, se puede construir un modelo matemático que defina el tiempo de solidificación de un lingote de acero, la velocidad de enfriamiento para evitar rajaduras de piezas de colada continua, un modelo para optimizar la recuperación metálica de una planta concentradora, etc.,

Los modelos matemáticos se clasifican en: teóricos, semi-empiricos, empíricos.

6.8.1 Modelos matemáticos teóricos

Son derivados fundamentalmente de las leyes físicas y se basan en la aplicación de ecuaciones de balance de materia y energía y en leyes de la termodinámica, cinética, etc. Los fenómenos de transferencia de calor y masa, flujos de fluidos, están dentro de esta categoría. Usualmente esas representaciones aparecen en forma de ecuaciones diferenciales parciales.

6.8.2 Modelos matemáticos semi-empíricos.

También están basados en las leyes físicas, pero se incorporan cierta cantidad de empirismo debido básicamente a un modelo muy complejo o la falta de datos.

6.8.3 Modelos matemáticos empíricos

Estos modelos no están basados en las leyes físicas. Se usan estos modelos cuando el proceso presenta una relación matemática muy compleja o desconocida y depende de las variables de entrada y salida, ignorando la estructura interior o fenómeno del proceso. La manera de construir este tipo de modelos es a través de las técnicas de *Diseños Experimentales*.

6.9 Experimentos con un solo factor

Frecuentemente en la experimentación se quiere comparar más de dos tratamientos. Por ejemplo, se quiere estudiar la efectividad de varios inhibidores de corrosión, el efecto de varios catalizadores en el rendimiento químico de un proceso, la durabilidad de varios tipos de material, etc. Los experimentos deberán ser diseñados para investigar los posibles efectos de tratamiento.

Se debe aquí enfatizar en el concepto de *aleatorización.* La aleatorización garantiza la validez de la inferencia frente a distorsiones no especificadas; hace que esas distorsiones se difundan uniformemente entre los tratamientos.

En los experimentos con un solo factor se debe efectuar experimentos completamente aleatorizados y se extiende el análisis de dos a *k* niveles (tratamientos) de un solo factor. En el análisis se asume que las muestras que se observan de los *k* tratamientos provienen de grupos independientes. En este diseño experimental las unidades son agrupadas en bloques homogéneos y los *k* tratamientos son aleatoriamente asignados a cada bloque.

6.9.1 Experimentos de un solo factor completamente aleatorizados

Asúmase que se tiene bajo estudio *k* tratamientos. Para el *i-ésimo* tratamiento, la respuesta *Y* es una variable aleatoria que varía alrededor de un promedio de tratamiento desconocido μ_i = 1, 2, ..., *k*; se asume que para cada grupo de tratamiento, la distribución de la respuesta *Y* alrededor del su promedio de grupo es *normal* y que las *varianzas son las mismas* para todos los grupos de tratamiento. Esta última asunción indica que la "*precisión"* de las observaciones es la misma para cada grupo. Un ejemplo, con *k*=3 distribuciones de tratamiento, se muestra en la Figura 6.4. Se asume que estas distribuciones obtenidas con muestras de diferentes grupos de tratamiento son independientes entre si, lo cual es muy importante para el análisis de datos.

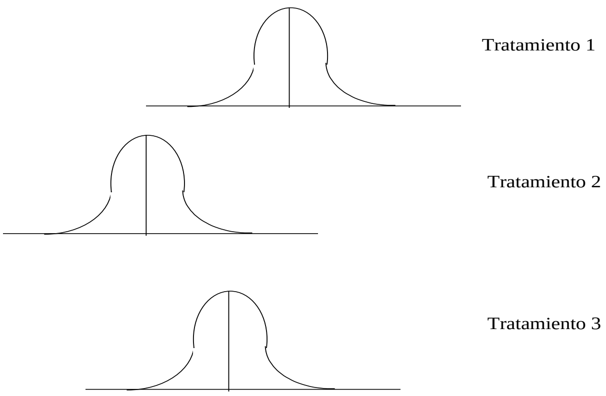

Figura 6.4: Gráfico de tres distribuciones normales con varianzas iguales.

Supóngase por ejemplo que se quiere estudiar la tensión de ruptura de tres diferentes aleaciones. Las mediciones deberán estar distribuidas normalmente y con varianza semejante. Se quiere averiguar si son iguales o no los promedios μ_1, μ_2,... μ_k de los *k* tratamientos. Si son iguales, se dice que no hay diferencia debida a los diferentes niveles de este factor. Si se determina que son diferentes, se necesitará determinar cuan diferentes son esos promedios entre si; así por ejemplo, pueden ser diferentes porque μ_1 es mayor que $\mu_2 = \mu_3$, o tal vez porque los tres son diferentes entre si.

6.9.2 Ejemplo de aplicación

Un ingeniero desea comparar las propiedades de resistencia de tres tipos diferentes de vigas. La de tipo *A* esta hecha de acero y las de tipo *B* y *C* son hechas de dos diferentes tipos de aleación más caras. Los resultados de deflección por aplicación de una fuerza de 3000 lb se anota en la Tabla 6.3. Para realizar los experimentos de medición, el ingeniero se aseguró que las vigas son muestras totalmente representativas, y no simplemente una elección de las primeras vigas que se tuvieron a la mano. Para asegurar la aleatoriedad de las mediciones, el ingeniero asignó un número al azar a cada una de las 20 vigas, y según ello realizó los experimentos.

Tabla 6.3: Deflexión de tres tipos diferentes de vigas

Tipo	n_i	*Observaciones*	$\overline{y_i}$	$\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}\right)^2$
A	8	82, 86, 79, 83, 85, 84, 86, 87	84	48
B	6	74, 82, 78, 75, 76, 77	77	40
C	6	79, 79, 77, 78, 82, 79	79	14

Una visualización clara de los resultados se muestra en el siguiente diagrama de puntos, Figura 6.5:

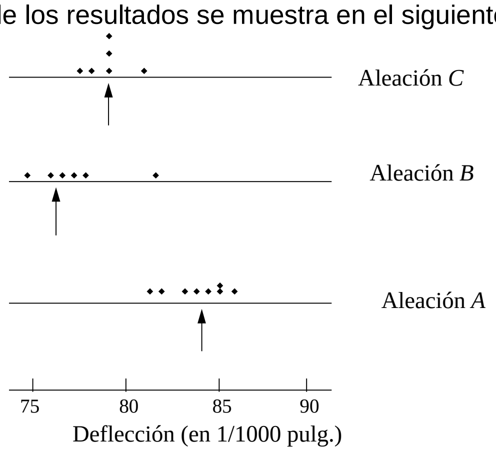

Figura 6.5: Diagrama de punto para los datos de tres grupos. El promedio de grupo se señala por la flecha.

El diagrama de puntos muestra dos aspectos muy importantes: a) La variabilidad de las observaciones dentro del grupo, o *variabilidad dentro del tratamiento*; b) La variabilidad entre los grupos, o *variabilidad entre tratamientos*.

Si la variabilidad entre tratamientos es más grande de lo que es de esperar, la posibilidad de que los tratamientos sean iguales queda seriamente cuestionada. El diagrama de puntos anotado, permite visualizar esta variabilidad entre grupos.

De modo análogo a la expresión de la varianza de muestra, se determina:

$$SS_{Error} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y_i}\right)^2$$

que es denominada *suma de cuadrados del error dentro del tratamiento* o *error de suma de cuadrados* (***E**rror **S**um **S**quares*). Esta cantidad, dividida por el número de grados de libertad, *N - k*, se conoce como *promedio del cuadrado del error* (***M**ean **S**quare **E**rror*) ó *promedio del cuadrado debido al error*:

$$MS_{Error} = \frac{SS_{Error}}{N-k}$$

Este es un promedio ponderado de la varianzas de muestra:

$$S_i^2 : \ MS_{Error} = \sum_{i=1}^{k} w_i S_i^2$$

donde $w_i = (n_i - 1)/(N-k)$. Este es un estimador insesgado de la varianza σ^2 y **mide la variabilidad interna**.

En el ejemplo, la suma de cuadrados dentro del tratamiento:

$SS_{Error} = 48 + 40 + 14 = 102$

y los grados de libertad están dados por: (8-1) + (6-1) + (6-1) = 17; por lo tanto el estimado insesgado de σ^2 es:

$MS_{Error} = 102/17 = 6.0$

Ahora, hay otro tipo de variabilidad que puede ser calculada; esto es, la variación entre los promedios de tratamiento alrededor de el gran promedio. El gran promedio:

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$$

es la suma de todas las observaciones dividida el número total de observaciones. Si a cada promedio $\overline{Y}_i$ se le da su tamaño de muestra n_i como su peso al calcular la suma de cuadrados, tal suma de cuadrados:

$$\begin{aligned} SS_{Tratamiento} &= n_i\left(\overline{Y}_1 - \overline{Y}\right)^2 + n_2\left(\overline{Y}_2 - \overline{Y}\right)^2 + \ldots + n_k\left(\overline{Y}_k - \overline{Y}\right)^2 \\ &= \sum_{i=1}^{k} n_i\left(\overline{Y}_i - \overline{Y}\right)^2 \end{aligned}$$

se denomina *suma de cuadrados entre tratamientos*, o simplemente *suma de cuadrados de tratamiento* (**T**reatment **S**um of **S**quares). Esta mide la variabilidad entre promedios de tratamientos. Será cero, si los promedios de tratamiento son iguales.

Intuitivamente se observa que esta suma de cuadrados tiene *k-1* grados de libertad, y a la proporción:

$$MS_{Tratamiento} = \frac{SS_{Tratamiento}}{k-1}$$

es el *promedio de cuadrados debido al tratamiento* (**M**ean **S**quare **T**reatment)

En el ejemplo: $SS_{Tratamiento} = 8(84 - 80.4)^2 + 6(77 - 80.4)^2 + 6(79 - 80.4)^2 = 184.8$

Ya que k - 1 = 3 - 1 = 2 grados de libertad, el promedio de cuadrados debido a tratamiento es:

$MS_{Tratamiento} = 184.8/2 = 92.4$

Definidos los dos tipos de variación: la suma de cuadrados dentro y entre tratamiento; la suma de cuadrados alrededor del gran promedio, *SSTO*, es la suma de estas dos sumas de cuadrados:

$SSTO = SS_{Error} + SS_{Tratamiento}$

7.9.2.1. <u>Tabla de Análisis de la Varianza</u>

Toda esa descomposición de sumas de cuadrados se registra frecuentemente en una Tabla. Ya que esta Tabla considera varias sumas de cuadrados o estimados de las varianzas, se le denomina *Tabla de Análisis de la Varianza* o Tabla *ANAVA* y se muestra en la Tabla 6.4. En esta Tabla, la primera columna identifica las fuentes de variación, la segunda las sumas de los cuadrados, la tercera los grados de libertad, la cuarta el promedio de los cuadrados y la última columna la proporción *F* que se analizara más adelante en esta sección.

Tabla 6.4: Tabla ANAVA para un diseño completamente aleatorizado.

Fuente	*SS*	*g.l.*	*MS*	*F*
Tratamiento	$\sum_{i=1}^{k} n_i\left(\overline{Y}_i - \overline{Y}\right)^2$	*k - 1*	$\frac{SS_{Tratamiento}}{k-1}$	$\frac{MS_{Tratamiento}}{MS_{Error}}$
Error	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2$	*N - k*	$\frac{SS_{Error}}{N-k}$	
Total	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}\right)^2$	***N - 1***		

En el ejemplo se tiene:

Tabla 6.5: Tabla ANAVA para el ejemplo de la vigas tipo *A, B* y *C.*

Fuente	*SS*	*g.l.*	*MS*	*F*
Tratamiento	184.8	*2*	92.4	15.4
Error	102.0	*17*	6.0	
Total	286.8	*19*		

6.9.2.2 Prueba *F* para los efectos de tratamiento

Se desea probar la hipótesis nula de que todos los tratamientos son iguales, versus la hipótesis alternativa de que todos los tratamiento son diferentes:

$H_o:\ \mu_1 = \mu_2 = ...\mu_n$

H_1 : Todos los tratamientos no son los mismos.

La prueba estadística para evaluar esto lo proporciona la razón $F = MS_{Tratamiento}/MS_{Error}$. El numerador, y por lo tanto esta razón es "pequeña" si H_o es cierta, lo que corresponde al caso en que los promedios de tratamiento son muy parecidos. Por otro lado, si el numerador es grande y por ende la razón señalada si H_1 es cierta, lo que corresponde al caso en que los promedios de tratamiento son diferentes.

La razón $MS_{Tratamiento}/MS_{Error}$. tiene distribución *F*, con *k - 1* y *N - k* grados de libertad. Si alguno de los promedios, μ_1, μ_2,... μ_k,es diferente, el numerador y la razón misma tienden a ser grandes. Esto conduce al siguiente procedimiento:

Si $F \geq F(\alpha; k - 1, N - k)$ se acepta H_1

En este caso se dice que la razón $F = MS_{Tratamiento}/MS_{Error}$. es significativa y que hay diferencias estadísticamente significativas entre los promedios de grupos. Por otro lado:

Si $F \leq F(\alpha; k - 1, N - k)$ se acepta H_o

En el ejemplo,. $F = MS_{Tratamiento}/MS_{Error}$. = 15.4. Para el valor crítico α = 0.05, de Tablas se obtiene *F*(0.05; 2. 17) = 3.59. Se concluye que existe fuerte evidencia de que los promedios de tratamiento son diferentes, de modo que se rechaza $H_o:\ \mu_1 = \mu_2 = ..., \mu_n$ a un nivel de significación de 0.05. Para el valor crítico α = 0.01, *F*(0.01; 2. 17) = 6.11, por lo cual también se rechaza $H_o:\ \mu_1 = \mu_2 = ...\mu_n$ a un nivel de significación de 0.01.

6.9.2.3 Procedimiento de Tukey para comparación múltiple

Una vez que se ha concluido que los promedios de tratamiento no son los mismos, se debe investigar de que forma difieren entre si. Por ejemplo, en el caso de la vigas, las del grupo *A* pueden ser diferentes a las del grupo *B* y *C* juntas, o que los tres tipos de vigas pueden ser diferentes entre si. Para obtener conclusiones valederas al respecto, se puede hacer comparaciones por pares. Si hay *k* tratamientos, se puede establecer *k(k*-1)/2 posibles tratamientos por pares. El error que se puede tener al hacer múltiples comparaciones se hace mayor que al hacer comparación de un solo par. Esto se tiene en cuenta cuando se establece el procedimiento de Tukey para comparación múltiple. Se calcula intervalos de confianza $\mu_\rho - \mu_s$ según:

$$\left(\overline{Y}_r - \overline{Y}_s\right) \pm \frac{q(\alpha;k,\nu)}{\sqrt{2}}\sqrt{MS_{Error}}\sqrt{\frac{1}{n_r}+\frac{1}{n_s}}$$

donde $q(\alpha;\ k,\ \nu)$ es el 100(1-α) por ciento de la *distribución Estudiantizada* (Tabla C8) que compara los promedios *k* con $\nu = N - k$ grados de libertad del promedio cuadrado de error. Esto permite establecer intervalos de confianza en comparaciones por pares. Si el intervalo que se defina para cada par de comparación incluye el cero, se concluye que no hay diferencia significativa entre los dos tratamientos. Estrictamente hablando, este procedimiento es válido solo para muestras de igual tamaño; sin embargo, es una buena aproximación si los valores n_i no son muy diferentes.

En el ejemplo de la comparaciones de la vigas, se puede establecer un total de 3(3-1)/2 = 3 comparaciones en pares, asociadas según:

$\mu_A - \mu_B$; $\mu_A - \mu_C$; $\mu_B - \mu_C$.

Ya que $q(0.05; 3, 17) = 3.62$, y :

$$\left[q(0.05;3,17)/\sqrt{2}\sqrt{MS_{Error}}\right] = 2.56\sqrt{6.0} = 6.27$$

los intervalos de la comparación Tukey son:

$$(84-77) \pm 6.27\sqrt{\frac{1}{8}+\frac{1}{6}}\ ;\ \text{o}\ 3.61 \le \mu_A - \mu_B \le 10.39$$

$$(84-79) \pm 6.27\sqrt{\frac{1}{8}+\frac{1}{6}}\ ;\ o\ \ 1.61 \le \mu_A - \mu_C \le 8.39$$

$$(79-77) \pm 6.27\sqrt{\frac{1}{6}+\frac{1}{6}}\ ;\ o\ -1.62 \le \mu_C - \mu_B \le 5.62$$

De esta comparaciones múltiples, se concluye que μ_A es mayor que μ_C y μ_B; y que μ_C , μ_B no son muy diferentes entre si.

Ejercicio de aplicación

La resistencia a la tracción de una fibra sintética es de interés de un fabricante. Se sospecha que la tensión es afectada por el porcentaje de algodón en la fibra. Cinco niveles en porcentaje de algodón son considerados y se efectúa 5 mediciones de cada nivel. Las 25 pruebas ejecutadas aleatoriamente se reportan a continuación.

% de Algodón	*Resistencia a la tracción, lbs/pulg^2*				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

a) Representar los datos usando gráficos de caja.

b) Construir la Tabla de ANAVA y probar la hipótesis: H_o: $\mu_1 = \mu_2 = \ldots = \mu_5$, H_1: al menos uno de los μ_i no es igual a los otros.

c) Graficar los promedios de tratamiento vs. El porcentaje de algodón utilizado.

d) Usar el procedimiento de Tukey para conducir una comparación múltiple.

e) Decidir si existe diferencia en la resistencia a la tracción debida a la cantidad de algodón utilizado.