

CAPÍTULO 4

MARCO TEÓRICO

Cabe mencionar que durante el proceso de medición, la precisión y la exactitud de cualquier magnitud física está limitada. Esta limitación se debe a que las mediciones físicas siempre contienen errores.

En este estudio, se espera encontrar si existe una diferencia significativa entre las dimensiones de una pieza dependiendo de su posición. Por tal motivo, para comprobar dicha diferencia de mediciones se utilizará el análisis de varianza. En este capítulo se encuentra la definición y el procedimiento de la ANOVA así como también el procedimiento de Tukey.

4.1 DEFINICIÓN DE ERROR E INCERTIDUMBRE

Para poder obtener una medición confiable, hay que controlar dichos factores causantes del error y conocer el nivel de incertidumbre del sistema de medición. Antes de continuar, es necesario definir la diferencia entre incertidumbre y error.

E. R. Cohen del Rockwell International Science Center, define error como la diferencia entre el valor medido de una cantidad física y el valor “verdadero” de esa cantidad. Es decir, es la desviación del valor real contra el valor nominal del objeto medido.

Define incertidumbre como la cantidad que describe la probable o posible magnitud del error desconocido. En otras palabras, incertidumbre se refiere al rango estimado donde se encuentra el valor verdadero medido.

Generalmente, los datos obtenidos de una medición, no son utilizados directamente como resultados, esta información requiere un ajuste o corrección. Esta corrección de los datos observados puede representarse como una función matemática dada por:

$$y = f(x_i; c_j; p_k) + \delta f$$

donde:

x_i = datos observados;

c_j = constantes: calibración, h, aceleración de la gravedad, etc.

p_k = parámetros: temperatura, presión, distorsiones mecánicas, etc.

δf = corrección de los datos analizados.

Los datos medidos son transformados en magnitudes físicas a través de la función f . La precisión de los resultados no sólo depende de la exactitud de la medición de los datos observados (x_i), si no que también involucra las constantes de calibración y la evaluación de los efectos de los parámetros ambientales.

Existen dos clasificaciones de errores experimentales: aleatorios y sistemáticos. Los errores aleatorios son asignados a la variable x_i , mientras que, los sistemáticos, a las variables c_j y p_k .

El proceso de reducción y corrección de los datos medidos es $y = f(x_i; c_j; p_k)$.
 Asumiendo que cualquier desviación del valor real fuese mínima, el error del valor de salida (y) está definido por:

$$\varepsilon_y = \sum_i \frac{\partial f}{\partial x_i} \varepsilon_{x_i} + \sum_j \frac{\partial f}{\partial c_j} \varepsilon_{c_j} + \sum_k \frac{\partial f}{\partial p_k} \varepsilon_{p_k} - \delta f$$

donde:

ε_{x_i} , ε_{c_j} y ε_{p_k} = error desconocido de los datos

Los errores ε definidos en la ecuación anterior, son desconocidos. Los valores asignados a x_i , c_j , p_k son estimaciones precisas y cualquier error conocido es automáticamente ajustado por la función f .

En un grupo de observaciones, $x_i = x_{\text{real}} + \varepsilon_i$, el error ε_i se puede descomponer en dos componentes, aleatorio y bias “b”. El error aleatorio tiene una varianza finita y una media cero. Mientras que, el error bias el cual es constante entre medición y medición, por lo que tiene media finita y varianza cero. Aunque no hay manera de determinar el error bias, solamente se puede obtener el valor $x_{\text{real}} + b$.

A parte de que no hay manera de separar el error b del valor real, también es imposible hacer una diferencia entre este error y el aleatorio, debido a que no es posible hacer mediciones iguales ya que las condiciones y parámetros no se mantienen constantes. Los componentes constantes que ocasionan el error b se refieren a los errores sistemáticos. Para conocer la incertidumbre sistemática es necesario modificar los factores ambientales de tal modo que se pueda determinar la forma en que afectan a los datos o mediciones. Con la aleatorización de errores sistemáticos se busca identificar la magnitud de los efectos y hacer la transformación de errores sistemáticos a correcciones conocidas. Los errores sistemáticos restantes son considerados como la incertidumbre de la magnitud del efecto.

Cabe mencionar que existe una diferencia entre errores sistemáticos e incertidumbre sistemática. El International Vocabulary of Metrology define error sistemático como “error que permanece constante o variable de forma predecible”. La parte predecible del error se puede corregir, sólo aquella parte residual del error sistemático conduce a una incertidumbre sistemática. Es decir, la incertidumbre sistemática se refiere a la parte del error sistemático no predecible. El término incertidumbre sistemática sólo se aplica a la contribución de errores b no corregidos o no medidos.

4.2 ANÁLISIS DE VARIANZA. ANOVA

Análisis de Varianza se puede definir como un conjunto de situaciones experimentales y procedimientos estadísticos para el análisis de respuestas cuantitativas de unidades experimentales (Devore, 2001). Los problemas de ANOVA, dependiendo del número de factores a estudiar, se clasifican en ANOVA de un solo factor y ANOVA con factores múltiples.

Los factores son las variables independientes que se pueden controlar. Estos factores afectan directamente la respuesta Y de un experimento. Durante el análisis de datos obtenidos durante un experimento, es necesario identificar las variables independientes o factores del experimento. Estas pueden ser de gran ayuda para la formación de bloques.

Para el análisis de varianza se emplea la prueba F para probar la hipótesis nula, la cual afirma la igualdad entre las medias poblacionales o las de tratamiento.

ANOVA de un solo factor hace una comparación de más de dos medias poblacionales o de tratamiento. Entonces la hipótesis nula es:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_l$$

contra:

H_a : al menos dos μ_i son diferentes

donde:

μ_1 = media de la población 1 o respuesta promedio del tratamiento 1

.

.

.

μ_I = media de la población I o respuesta promedio del tratamiento I

I = número de poblaciones o tratamientos en comparación

La hipótesis nula, H_0 , es verdadera, entonces las observaciones de cada muestra tienen como origen una distribución normal con la misma media, μ . Por lo tanto, las medias muestrales deben ser similares. El estadístico de prueba está basado en comparar la variación entre muestras, con una medida de variación calculada a partir de cada una de las muestras.

El estadístico de prueba para una ANOVA de un solo factor se define como:

$$F = \frac{MSTr}{MSE}$$

donde:

MSTr = Cuadrado Medio de Tratamientos

MSE = Cuadrado Medio del Error

4.3 DISTRIBUCIÓN “F”

La distribución F tiene dos parámetros enteros positivos, v_1 y v_2 . El primero se conoce como número de grados de libertad del numerador y el segundo, v_2 , número de grados de libertad del denominador. La distribución F se relaciona fuertemente con la distribución χ^2 . Si X_1 y X_2 son variables aleatorias χ^2 independientes entre sí con v_1 y v_2 grados de libertad, entonces se demuestra que:

$$F = \frac{X_1/v_1}{X_2/v_2}$$

La notación de la esta distribución es F_{α, v_1, v_2} . La curva de densidad de la distribución F no es simétrica por lo que tenemos:

$$F_{1-\alpha, v_1, v_2} = \frac{1}{F_{\alpha, v_1, v_2}}$$

4.4 PRUEBA “F”

Teorema:

Sea $F = \text{MSTr}/\text{MSE}$ el estadístico de prueba de un análisis de varianza de un solo factor, con “k” poblaciones o tratamientos y una muestra aleatoria de “n” observaciones. Si H_0 es verdadera, entonces F tiene una distribución F con k-1 grados de libertad en el numerador y k(n-1) grados de libertad en el denominador. Si f representa el valor calculado de F, la región de rechazo $f \geq F_{\alpha, k-1, k(n-1)}$ especifica una prueba con nivel de significancia α .

4.5 PROCEDIMIENTO DE TUKEY

Cuando H_0 es rechazada, generalmente se requerirá saber cuales medias poblacionales o de tratamientos son diferentes. Para poder identificar las medias diferentes, existen varios métodos llamados procedimientos de comparaciones múltiples.

Para este procedimiento se requiere la intervención de la distribución de rango estudentizado. Esta distribución requiere de dos parámetros: grados de libertad m del numerador y grados de libertad v del denominador. Entonces tenemos que $Q_{\alpha,m,v}$ es el valor crítico de cola superior α de la distribución. Con el valor crítico $Q_{\alpha,I,I(J-1)}$ se puede utilizar para generar intervalos de confianza simultáneos para las diferencias por pares de las medias de las diferentes poblaciones o tratamientos ($\mu_i - \mu_j$).

Con la probabilidad de $1-\alpha$, tenemos:

$$\bar{X}_{i\bullet} - \bar{X}_{j\bullet} - w \leq \mu_i - \mu_j \leq \bar{X}_{i\bullet} - \bar{X}_{j\bullet} + w$$
$$w = Q_{\alpha,I,I(J-1)} \sqrt{\frac{MSE}{J}}$$

para toda i y j , con $i \neq j$.

Con la función anterior, se obtiene una colección de enunciados simultáneos de confianza sobre los valores reales de las diferencias entre $\mu_i - \mu_j$, entre las medias reales de la población o tratamiento. Es decir, cada intervalo que no incluya cero nos lleva a la conclusión de que hay una diferencia significativa a nivel α entre los valores de μ_i y μ_j .