

# Proyecto 3: Inferencia estadística

Torres G. Lesmes      Barrios S. Xavier      Calderon G. Jose Andres  
Tecnológico de Costa Rica      Tecnológico de Costa Rica      Tecnológico de Costa Rica  
Correo: lesmestorres@estudiantec.cr      Correo: xbs20@estudiantec.cr      Correo: jandres2019@estudiantec.cr

**Resumen**—En este informe, se muestra la comparación de resultados sobre el análisis de una máquina manufacturera en una empresa, en la cual se busca mejorar su rendimiento. Se realizaron 3 cambios en la configuración de la máquina, por lo que se espera que aumente el rendimiento promedio requerido en un 5 por ciento. Mediante el uso de la inferencia estadística y otros análisis, se concluye que la configuración inicial es la más beneficiosa para la empresa.

**Index Terms**—Análisis, Estimación, Hipótesis, Parámetros, Rendimiento.

## I. INTRODUCCIÓN

En una empresa se tiene una máquina manufacturera, la cual debe trabajar bajo un rendimiento de operación de al menos un 70 % para considerarse aceptable. La máquina opera 24 horas cada día de la semana y su rendimiento es medido diariamente. Un ingeniero realizó cambios en la configuración de la máquina para mejorar el rendimiento promedio requerido y este asegura un incremento del 5 % sobre rendimiento promedio requerido. Esta conclusión se fundamenta en las mediciones diarias durante una semana, con un rendimiento promedio de 75 %. Por lo cual, al respaldarse de una muestra pequeña de 7, genera dudas respecto a su consistencia, por lo que mediante distintas técnicas de muestreo, prueba e hipótesis, inferencia estadística y estimación de parámetros, se estudian 3 escenarios distintos, durante 90 días, se registró el rendimiento de tres configuraciones diferentes: una con la configuración del colega (primer cambio), otra con una configuración propuesta (segundo cambio) y un registro de medición de la máquina antes de cualquier implementación de configuración nueva (inicial). De estas tres configuraciones, mediante análisis, se justificó a la mejor en términos de rendimiento, con el objetivo de obtener el mayor beneficio para la producción de la empresa.

## II. JUSTIFICACIÓN

Para el estudio de las tres configuraciones de la máquina, primeramente se graficaron los datos recopilados mediante un histograma, el cual exhibe un comportamiento similar a una distribución normal. Mediante el método de aproximación de la normal se logró visualizar este comportamiento para los datos recopilados.

Posteriormente, con los datos de medición recopilados de las tres configuraciones de la máquina, se determinó

la media, la desviación estándar y la varianza para cada una de ellas. Esto con el objetivo de corroborar que cumplieran con el requisito mínimo del rendimiento del 70 % en su rendimiento medio, además de conocer la dispersión de los datos en las mediciones tomadas. Debido a que las medidas determinadas de tendencia central y dispersión pertenecen a una muestra de 90 mediciones en cada configuración y se desea conocer el comportamiento real de la máquina. Se utiliza el método de máxima verosimilitud (MLE) para encontrar el estimador de los parámetros para la media, varianza y la desviación estándar con error de estimador intrínseco. Estos estimadores MLE son bastante conocidos para una distribución normal, estos se resumen en Tabla I.

Tabla I

Estimadores utilizados		
Parámetro	Muestral	MLE
Promedio	$\mu$	$\mu$
Desviación estándar	$\sigma$	$\frac{\sigma}{\sqrt{n}}$
Varianza	$\sigma^2$	$\frac{\sigma^2}{n}$

*Nota:* Los estimadores MLE representan los estimadores de máxima verosimilitud.  $\mu$  es el promedio muestral,  $\sigma$  es la desviación estándar muestral y  $\sigma^2$  es la varianza muestral. El símbolo  $n$  representa el tamaño de la muestra [1].

El valor de la varianza de la muestra, difiere respecto al valor de la varianza proveniente del método de máxima verosimilitud (MLE), esta diferencia se calcula con un porcentaje de error. Esto permite manejar un valor cuantitativo del porqué se debe utilizar un estimador proveniente del método de máxima verosimilitud.

Se va a suponer que el valor de media muestral, determinado de forma analítica, es el valor real conocido para la distribución estudiada en cada configuración, de manera que la desviación estándar y varianza pueden usar este valor de media para determinarse. Como nota importante, en los posibles estimadores mostrados en la Tabla I, se ve que la media para un estimador muestral y el estimador determinado por método de máxima verosimilitud (MLE) es el mismo, esto justificaría el uso de la media muestral como valor conocido.

La inferencia estadística la cual se maneja con la

prueba de hipótesis, en esta se plantean dos supuestos, una hipótesis nula y su alternativa en caso que la primera sea falsa, para un parámetro de una población [1].

En cada caso de prueba de hipótesis, se tienen dos hipótesis, la nula que se representa como  $H_0 : \mu = \mu_0$  y  $H_1 : \mu >, < o \neq \mu_0$  que sería la hipótesis alternativa [1]. La hipótesis nula siempre busca rechazarse con la prueba de hipótesis, su rechazo lleva siempre a aceptar la hipótesis alternativa.

Ahora, para este método se tiene siempre un porcentaje de error e incertidumbres asociadas, lo cual nos lleva al tipo de error tipo I, donde se rechaza la hipótesis nula, aunque ésta sea verdadera. En el caso de las configuraciones de las máquinas a analizar es conveniente hacer la prueba de hipótesis, para obtener los parámetros y llegar a la inferencia estadística.

Para la configuración de primer cambio mostrada en la Figura 2, se asegura que existe un 5 % de mejora sobre el rendimiento mínimo aceptable de 70 %. Para comprobar esta afirmación se procede a evaluar como una probabilidad de Error de tipo 1, que representa probabilidad del error de rechazar la hipótesis nula siendo esta verdadera, esta se conoce como  $\alpha$ .

Para evaluar el 5 % de mejora en el rendimiento, tanto para la prueba del primer cambio como de las otras dos, se utiliza como este valor de 5 % como  $\alpha$ , este debe ser corroborado mediante el valor P, que mide el nivel de significación más bajo para rechazar la hipótesis nula [1]. Debido a que no se cuenta con un valor conocido de varianza, se debe determinar el valor P, utilizando la prueba T.

La prueba T permite medir la diferencia entre la media muestral y la media poblacional en términos de la variabilidad dentro de la muestra, en el caso de las configuraciones de la máquina, se cuenta con un valor de media muestral y desviación estándar. La hipótesis nula defiende que hay poca diferencia entre el valor de la media para la muestra y de la población.

El valor T, tiene un valor  $\alpha$  asociado, como la hipótesis nula defiende que la diferencia entre las medias o valor T, muestral y poblacional deben ser la cero o al menos muy baja, esta cuenta con un valor crítico de T asociado al valor  $\alpha$ , si el T calculado por el estadístico de la distribución t, supera el t crítico, se puede rechazar la hipótesis nula [1]. Y para cada  $\alpha$  existe un valor P que se puede calcular.

Al incluir en la prueba T dos muestras, se puede comparar la diferencia existente entre la media de cada muestra y la significancia que tiene una respecto a la otra [1]. El valor T que arroja dicha prueba se compara con el rango de valores específicos para el valor de  $\alpha$  seleccionado y los grados de libertad, con lo cual se tiene que si el valor obtenido por la prueba es muy cercano a los valores en el intervalo esto quiere decir que la diferencia entre las medias de cada muestra

es poco significativa. Por otro lado el valor P indica la probabilidad de obtener una diferencia significativa entre las medias de las dos muestras.

### III. ANÁLISIS DE RESULTADOS

En la Figura 1, Figura 2 y Figura 3, se observa la distribución de los datos de las muestras Inicial, Primer cambio y Segundo cambio respectivamente, en las cuales se puede observar que los datos poseen un comportamiento normal y seguidamente se aproximan a la distribución normal, con la cual se estimó la varianza y la desviación estándar. En cada caso se tomó el valor de la media muestral determinada de forma analítica como el valor de la media para la distribución normal.

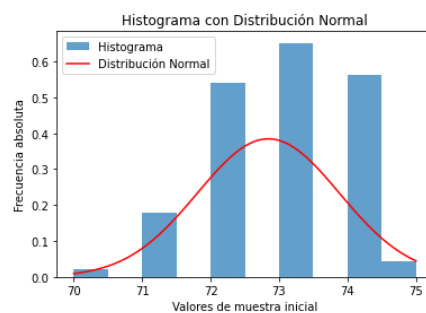


Figura 1: Gráfica de la distribución de la configuración inicial

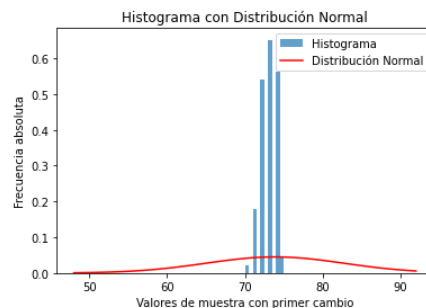


Figura 2: Gráfica de la distribución de la configuración con el primer cambio

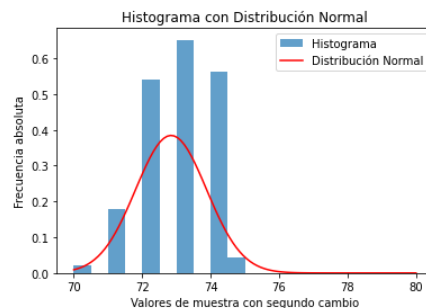


Figura 3: Gráfica de la distribución de la configuración con el segundo cambio

Como se puede observar en la Tabla III los valores de P son mas pequeños que el valor de  $\alpha$ , por lo cual en los tres casos existe la evidencia suficiente para rechazar la hipótesis nula. Con esto se puede decir que las tres configuraciones son aceptables.

Con respecto al valor T que se obtuvo de la prueba T, se tiene que estas configuraciones de la máquina son más grandes que el valor T crítico el cual se encuentra en el rango  $[1,658 - 1,671]$  para un  $\alpha$  del 5 % como se puede observar en la Tabla III, por lo tanto, en los tres casos se respalda el hecho de rechazar la hipótesis nula.

Tabla II

Medidas de tendencia central para las tres muestras			
Datos	Inicial	Primer cambio	Segundo cambio
Promedio	72,84 %	73,85 %	75,29 %
Desviación estandar muestral	1,04 %	8,94 %	2,13 %
Desviación estandar MLE	1,03 %	8,89 %	2,11 %
Varianza MLE	1,08	79,14	4,48
Porcentaje de error de la varianza estandar	1.12 %	1.12 %	1.12 %
Porcentaje de error de la desviación estandar	0,56 %	0,56 %	0,56 %

De forma general se puede observar en la Tabla III que la muestra de la configuración Inicial posee el valor T mas alto y valor P mas bajo de las tres muestras, por lo tanto es en la que se puede rechazar la hipótesis nula con mayor seguridad. Por otro lado la muestra generada por la configuración realizada por el colega ingeniero posee el valor P más alto y el valor T mas cercano al intervalo de valores críticos, por lo cual, esta muestra es la que posee el menor peso para rechazar la hipótesis nula.

Tabla III

Valores de la prueba T			
Datos	Inicial	Primer cambio	Segundo cambio
Valor P	5,52e-43	5,22e-05	6,12e-40
Valor T	25,70	4,06	23,46

Tabla IV

Comparación entre muestras			
Datos	Inicial vs primer cambio	Inicial vs segundo	Primer cambio vs segundo cambio
Valor P	0,29	3,87e-17	0,14
Valor T	-1,05	-9,75	-1,45

### III-A. Inferencia estadística

1. ¿Son todas las configuraciones para la maquina aceptables?

Se puede observar en la Tabla III que los valores de P son menores a 0,05 en todos los casos, por lo tanto se rechaza la hipótesis nula dando paso a la hipótesis alternativa, la cual indica que el rendimiento es mayor a 70.

2. ¿Es correcta la conclusión del colega ingeniero en el caso hipotetico respecto a la primera mejora? ¿Si, no, casi, y por que?

Al observar el valor T y P en la Tabla III se nota que los valores superan los limites necesarios para aceptar la hipótesis alternativa, por lo tanto, la solución es correcta.

3. ¿Cual de las configuraciones es la que mas beneficia a la empresa?

Segun los valores p de la tabla IV podemos ver que el valor P de las configuraciones inicial vs segundo cambio es muy pequeño, por lo que el porcentaje de error que hay entre los cambios de estas dos casi no varia. Además, debido a que la muestra de la configuración Inicial posee el valor T mas alto y valor P mas bajo de las ambas, es la en que se puede rechazar la hipótesis nula con mayor seguridad.

4. ¿Poseen las tres configuraciones la misma varianza? ¿En que afecta al resultado de comparación entre configuraciones posibles diferencias en las varianzas de las distribuciones?

Como se puede observar en la Tabla II las varianzas son diferentes para cada muestra, debido a que usan su propia media y esto afecta los resultados de los estadísticos (MLE) utilizados para determinar cada una de ellas. La varianza indica el grado de variabilidad de los datos y con esto se ve reflejado en la configuración de primer cambio, en este caso de necesitar estabilidad en el rendimiento de la máquina no podría utilizarse.

#### IV. CONCLUSIÓN

Se puede concluir que todas las configuraciones de las máquinas son aceptables, ya que en todas se rechaza la hipótesis nula, pero la mejor configuración a utilizar para aumentar el beneficio a la empresa es la configuración inicial.

#### REFERENCIAS

- [1] D. M. y G. Runger, *Probabilidad y estadística aplicadas a la ingeniería*.