UNIVERSIDAD DE SONORA
DIVISIÓN DE INGENIERÍA

POSGRADO EN INGENIERÍA INDUSTRIAL

MAESTRÍA EN INGENIERÍA EN SISTEMAS Y TECNOLOGÍA

ANÁLISIS ESTADÍSTICO DE DOS POBLACIONES MULTIVARIADAS PARA LA COMPARACIÓN DE GAMAS DE COLOR EN LA IMPRESIÓN FLEXOGRÁFICA DE ETIQUETAS Y EMPAQUES

T E S I S

PRESENTADA POR

GLORIA LUCERO RAMÍREZ ZAMORA

Desarrollada para cumplir con uno de los
requerimientos parciales para obtener
el grado de Maestra en Ingeniería

DIRECTOR DE TESIS
M.C. CARLOS ANAYA EREDIAS

CODIRECTOR
DR. JAIME OLEA MIRANDA

HERMOSILLO, SONORA, MÉXICO. FEBRERO 2022

UNIVERSIDAD DE SONORA

"El saber de mis hijos hará mi grandeza"

División de Ingeniería
Posgrado en Ingeniería Industrial
Maestría en Ingeniería en Sistemas y Tecnología

Hermosillo, Sonora a 13 de diciembre de 2021.

GLORIA LUCERO RAMIREZ ZAMORA

Con fundamento en el artículo 66, fracción III, del Reglamento de Estudios de Posgrado vigente, otorgamos a usted nuestra aprobación de la fase escrita del examen de grado, como requisito parcial para la obtención del Grado de Maestro(a) en Ingeniería: Ingeniería en Sistemas y Tecnología.

Por tal motivo este jurado extiende su autorización para que se proceda a la impresión final del documento de tesis: **ANÁLISIS ESTADÍSTICO DE DOS POBLACIONES MULTIVARIADAS PARA LA COMPARACIÓN DE GAMAS DE COLOR EN LA IMPRESIÓN FLEXOGRÁFICA DE ETIQUETAS Y EMPAQUES** y posteriormente efectuar la fase oral del examen de grado.

ATENTAMENTE

Carlos Anaya Eredias

M.C. CARLOS ANAYA EREDIAS
Director(a) de tesis y Presidente del jurado

M.C. GUILLERMO CUAMEA CRUZ
Secretario(a) del Jurado

DRA. RAQUEL TORRES PERALTA
Vocal del Jurado

DR. JAIME OLEA MIRANDA
Vocal del Jurado

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Santiago, Chile, a 24 de diciembre de 2021

GLORIA LUCERO RAMIREZ ZAMORA

Con fundamento en el artículo 66, fracción III, del Reglamento de Estudios de Posgrado de la Universidad de Sonora, otorgo a usted mi aprobación de la fase escrita del examen profesional, como requisito parcial para la obtención del Grado de Maestra en Ingeniería: Ingeniería en Sistemas y Tecnología.

Por tal motivo, como sinodal externo y vocal del jurado, extiendo mi autorización para que se proceda a la impresión final del documento de tesis: **ANÁLISIS ESTADÍSTICO DE DOS POBLACIONES MULTIVARIADAS PARA LA COMPARACIÓN DE GAMAS DE COLOR EN LA IMPRESIÓN FLEXOGRÁFICA DE ETIQUETAS Y EMPAQUES** y posteriormente efectuar la fase oral del examen de grado.

ATENTAMENTE

DR. JUAN DOMINGO VELÁSQUEZ SILVA
UNIVERSIDAD DE CHILE
Sinodal Externo y Vocal del Jurado

RESUMEN

Siendo el proceso de conversión más utilizado en la industria del embalaje, la flexografía tiene la función de producir productos como empaques, películas y etiquetas autoadheribles que, además de ser útiles para el resguardo de alimentos y bebidas e identificación de productos, deben de ser lo suficientemente atractivos para acaparar la atención de un consumidor final; esto, debido a que el empaque, película y etiqueta forman parte de la presentación del producto, mismos que se usan como herramientas de mercadotecnia para así lograr un mejor posicionamiento en el mercado. Por lo antes mencionado, la industria de conversión tiene la obligación de producir productos que cumplan con las especificaciones del cliente, lo que significa que, además de ser llamativos, deberán de presentar tonalidades y colores consistentes entre familia de productos que contengan mismo color de logotipo e imágenes representativas de una marca.

En este estudio se compararon, utilizando el estadístico T^2 de Hotelling, los vectores medios obtenidos de las tintas Cian, Magenta, Amarillo, Negro, Naranja, Verde y Violeta de dos diferentes proveedores y con dos diferentes tecnologías de secado (UV estándar y UV LED). Para ello se realizaron impresiones, en 2 máquinas diferentes, y sobre dos tipos de materiales, el primero de ellos fue un semigloss y el segundo un polipropileno biorientado; debido a la necesidad de imprimir productos que presenten misma tonalidad en colores e imágenes, se esperaría encontrar que no existe diferencia significativa entre la gama de color de las dos máquinas. Antes de realizar las pruebas, se buscó igualar en componentes ambas máquinas de impresión; una vez igualadas, se imprimieron 20 hojas por cada máquina y por cada proveedor. Con la ayuda de un espectrofotómetro se tomaron mediciones de las coordenadas L, a y b, así como la densidad de tinta sólida en cada una de las hojas impresas, obteniendo 4 variables respuesta por medición. Los resultados mostraron que a pesar de igualar en componentes e insumos ambas máquinas, la impresión de todos los colores en ambos proveedores, resultaron con vectores medios diferentes, significando que los parámetros de impresión son diferentes para ambas máquinas.

ABSTRACT

Being the most used conversion process in the packaging industry, flexography has the function of producing products such as packaging, films and self-adhesive labels that, in addition to being useful for the protection of food and beverages and product identification, must be as attractive enough to capture the attention of an end consumer; this, because the packaging, film and label are part of the product presentation, which are used as marketing tools to achieve a better position in the market. Due to the aforementioned, the converting industry has the obligation to produce products that meet the customer's specifications, which means that, in addition to being striking, it must present contrasting shades and colors among the family of products that contain the same color of logo and representative images of a brand.

In this study, using Hotelling's T^2 statistic, the mean vectors obtained from Cyan, Magenta, Yellow, Black, Orange, Green and Violet inks from two different suppliers and with two different drying technologies (standard UV and UV) were compared. ultraviolet LEDs). For this, impressions were made, in 2 different machines, and on two types of materials, the first of them was a semigloss and the second a bioriented polypropylene; Due to the need to print products that have the same tonality in colors and images, one would expect to find that there is no significant difference between the color gamut of the two machines. Before carrying out the tests, it was sought to match both printing machines in components; once matched, 20 sheets were printed for each machine and for each supplier. With the help of a spectrophotometer, measurements of the L, a and b coordinates were taken, as well as the density of solid ink in each of the printed sheets, obtaining 4 response variables per measurement. The results showed that despite equalizing components and supplies for both machines, the printing of all colors in both providers resulted in different media vectors, meaning that the printing parameters are different for both machines.

DEDICATORIA

A mis padres, Rosario y Héctor Mario, quienes siempre me han enseñado que como profesionista uno debe de seguir preparándose, aprendiendo y estudiando temas nuevos y no sólo quedarse con lo visto en la ‘carrera’.

AGRADECIMIENTOS

A mi director de tesis, el maestro en ciencias Carlos Anaya Eredias, principalmente por haber aceptado, desde el primer semestre, mi ante proyecto de tesis. Estoy sumamente agradecida por su paciencia, disposición y tutoría durante la realización de mi proyecto de tesis. Gracias por sus enseñanzas y explicaciones en los temas relacionados con estadística multivariada y que, aun siendo complicados, logró que yo los entendiera.

Al coordinador del posgrado, el doctor Alonso Pérez Soltero, por estar al pendiente, así como de la orientación brindada durante el proceso de culminación de tesis.

Al Ing. Stone por introducirme al mundo de la teoría del color, así como a los diseñadores gráficos Paola, Rubén y Laura, quienes me enseñaron cómo funciona el proceso de Pre-Prensa, Rubén, muchas gracias por ayudarme con el diseño de algunas imágenes que adjunto en el marco teórico. A los chicos del área de impresión quienes, además de enseñarme del proceso, me ayudaron con la impresión de las muestras utilizadas en este estudio.

A Gildardo, por apoyarme y motivarme.

Por último y no menos importante, al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al Programa de Fortalecimiento de la Calidad Educativa (PFCE) por su apoyo económico brindado en mi estudio de posgrado.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

1. INTRODUCCIÓN

El etiquetado y empaquetado de productos puede usarse como una herramienta de mercadotecnia. Para Creusen y Schoormans (2005) el empaque es una forma de atraer la atención del consumidor hacia el producto, así mismo, Silayoi y Speece (2007) mencionan que el empaque se comunica con los consumidores en el momento en el que están tomando decisiones en la tienda, convirtiéndose en un factor crítico para la toma de decisiones; de la misma forma éste brinda información, permitiéndole al cliente conocer, de forma general, el contenido del producto. Se sabe que, la mercadotecnia a lo largo de los años ha desarrollado técnicas que ayudan a las empresas a consolidarse en el mercado, en efecto, el diseño de logotipos con colores representativos da origen a una identidad de marca, misma que logra hacer que un producto destaque de entre la competencia. Ampuero y Vila (2006) consideran que el empaque es “el rostro más visible de la marca” y tiene la gran capacidad de crear diferenciación e identidad del producto. Por consiguiente, se tiene una necesidad de obtener etiquetas y empaques que sus colores sean consistentes y no presenten variación entre familias de productos o reimpresiones.

1.1 Presentación

El proyecto se realizará en una empresa dedicada a la impresión de etiquetas autoadheribles y películas plásticas, cuyo propósito es que sus clientes puedan etiquetar y/o empacar (en caso de película) sus productos y que éstos llamen la atención de un consumidor final.

La empresa cuenta con varios departamentos: gerencia, diseño, pre-prensa, impresión, corte, bolseo, calidad, inspección, ventas, mercadotecnia, sistemas, tintas, facturación, empaque, y mantenimiento. Los departamentos de pre-prensa e impresión llevan a cabo las actividades críticas del proceso de producción de etiquetas y empaques.

El área de pre-prensa lleva a cabo las actividades relacionadas con el diseño de la etiquetas y bolsas, así como la fabricación de placas. Una vez aprobado el diseño por el cliente, se crean las placas de impresión, hechas con fotopolímeros grabados en relieve. Para la producción de placas es necesario separar el diseño en capas de colores los cuales pueden ser cian, magenta, amarillo, negro, naranja, verde y/o violeta. Además, y de ser necesario, se agregan colores directos (también conocidos como institucionales ejemplo, el rojo *coca-cola)*, los cuales son utilizados para la impresión de un logotipo de marca. El número total de placas es definido por el total de tintas necesarias para lograr la impresión del diseño.

El área de impresión, también conocida como el área de prensa, actualmente está conformada por 6 máquinas. Es importante mencionar que, ninguna máquina (prensa) cuenta con una configuración totalmente igual a otra debido a que a cada una de ellas la componen un determinado número de estaciones y las cuales están conformadas por: lámpara de secado de tinta, charola, entintador, rodillo anilox, navaja y cilindro porta placa. Actualmente, en el área de impresión se tiene: una máquina con un total de 4 estaciones y 4 máquinas con un total de 8, de las 6 máquinas una de ellas tiene lámparas de secado LED, mientras que los 5 restantes tienen lámparas de mercurio. Cabe aclarar que en esta empresa se utilizan las tecnologías de secado de tinta UV estándar y UV LED.

Durante la impresión, la configuración de las máquinas se hace bajo las especificaciones del producto a imprimir, ya que el número de tintas que el producto detalla es igual al número de estaciones a utilizar en la máquina, así como el número de placas producidas por pre-prensa.

El proceso de impresión inicia cuando el impresor registra un color tras otro, haciendo una sobreimpresión de las tintas y teniendo como resultado final un tono, color y/o imagen. La amplia cantidad de componentes (*anilox, sticky back,* impresor, en origen de tintas, el proveedor de papel, por mencionar algunos) provocan que el tono de impresión no sea el mismo de un pedido a otro. Esta falta de consistencia genera molestia al cliente y, siendo considerado como un problema de calidad, se ha implementado el control de color en impresión; mismo que trata de una medición para cada una de las tintas impresas. La medición, realizada con un espectrofotómetro, obtiene valores numéricos (L, a, b y densidad), mismos que se comparan matemáticamente con los valores estándar de cada uno de los colores. En caso de estar fuera de especificación, es necesario hacer cambios en los componentes de la máquina; estos ajustes requieren hacer paros de máquina, ya que, para lograr un color o tono dentro de especificación, deberán cambiar la configuración de la máquina.

1.2 Planteamiento del problema

El proceso de impresión presenta una alta inconsistencia de colores impresos, esto debido posiblemente a la variación que presentan los proveedores de tintas entre ellos, la diferencia en configuración de cada una de las máquinas en el área de prensa, el sustrato de impresión, así como la experiencia y habilidad de cada uno de los impresores.

1.3 Objetivo general

Determinar, mediante la comparación de dos poblaciones multivariadas, si ambas máquinas de impresión presentan los mismos parámetros en la gama de color de impresión.

1.4 Objetivos específicos

En este trabajo se tienen como objetivo los siguientes puntos:

- Comparar la reproducción de color entre las máquinas “5” y “7”.
- Realizar un análisis estadístico multivariado con el fin de obtener el resultado del desempeño de la gama de color de cada máquina.
- Implementar y estandarizar el proceso, de ambas máquinas, a partir del resultado de la comparación de dos poblaciones multivariadas.

1.5 Hipótesis

Con el análisis de comparación de dos poblaciones multivariadas se espera encontrar que no existe diferencia significativa entre la gama de color de las dos máquinas.

1.6 Alcances y delimitaciones

Aun cuando el área de prensa cuenta con 5 máquinas de impresión, el estudio se llevará a cabo sólo en las máquinas “5” y “7”, puesto que son los equipos de trabajo con más demanda. Para la máquina “5” el estudio se hará con material BOPP (Polipropileno Biorientado), el cual por sus características es transparente y es necesario aplicarle un “fondeado” de tinta blanca. Para la “7” el estudio se hará en un sustrato llamado *semigloss* (papel blanco con acabado semi-brillante).

1.7 Justificación

En el último par de años la empresa ha buscado implementar nuevos procedimientos para evaluar la calidad, con el objetivo de reducir el número de las no conformidades atribuibles al color. Uno de los nuevos métodos de medición para el control de calidad en la impresión de etiquetas y empaques es la medición de color, este nuevo procedimiento tiene el propósito de medir, por medio de un espectrofotómetro, todas las impresiones de etiquetas y empaques que se imprimen día a día.

Llevar a cabo un estudio estadístico de comparación de dos poblaciones dará información a la empresa respecto al alcance que tiene cada máquina impresora en el gamut de color. Es importante mencionar que, si ambas poblaciones resultan ser iguales, el proceso de impresión se verá estandarizado, y de resultar lo contrario se deberán de hacer modificaciones en base a las conclusiones que se obtengan del estudio para así obtener, en algún futuro, una homogeneidad en el resultado de impresión.

2. MARCO DE REFERENCIA

En este capítulo se definen los temas que comprende este documento, tales como impresión, impresión flexográfica, medición del color, gamut de color, formulación de color en impresión, análisis multivariado y posteriormente se presentan trabajos previos. Esto con la finalidad de brindar al lector una amplia visión del contenido de este trabajo.

2.1 Impresión

La impresión, definida por la enciclopedia Rosato (2000), es un método utilizado para la decoración y ejecución de los objetivos de marketing, que se imponen a los materiales plásticos. La industria de la impresión de envases utiliza tradicionalmente colores de proceso CMYK para reproducir imágenes fotográficas y colores directos para reproducir grandes áreas sólidas e información crítica de la marca como logotipos. (Deshpande, 2015). Existen diferentes técnicas de impresión que permiten transformar un diseño a un empaque o etiqueta; éstas pueden ser por medio de procesos de impresión tales como la inyección de tinta, litografía, offset, huecograbado, serigrafía, sublimación, flexografía, solo por mencionar algunos. En la figura 2.1 se pueden observar todas las diferentes técnicas que existen.

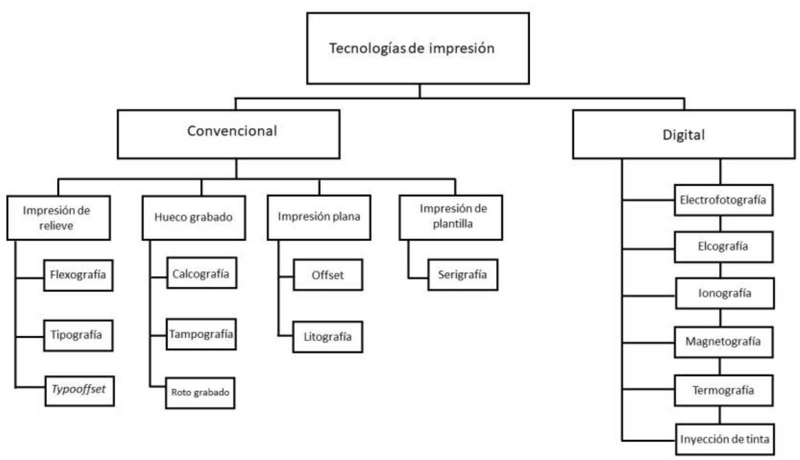

Figura 2.1 Técnicas de impresión.

2.2 Flexografía

Harris *et al.* (2004) la definen como “un método general y versátil de impresión a gran escala en papel (como, por ejemplo, en y para periódicos, hojas publicitarias), cartón, y sustratos de película plástica. Tal procedimiento implica la utilización de impresión de tipo relieve, que incluye caucho elevado y / o placas poliméricas con imágenes presentes en el mismo.” Así mismo, Patel (2009) reconoce a la impresión flexográfica como el proceso más ampliamente utilizado en la industria del embalaje. La diferencia de la impresión flexográfica con otros estilos es que permite imprimir a grandes volúmenes y en diferentes tipos de sustratos (mencionados en la definición anterior); Prust (2010) define sustrato como “cualquier material con una superficie que pueda imprimirse o recubrirse”. Estos materiales son suministrados en forma de rollo (bobinas) para introducirlos en máquinas de procesamiento continuo de bandas. En la figura 2.2 se observa que el material suministrado recorre, de manera continua, un total de 8 estaciones, mismas que representan el total de tintas que la máquina puede imprimir.

Figura 2.2 Impresora flexográfica.

2.3 Color

El color, definido por Hirschler *et al.* (2018) “es una característica de la percepción visual que puede describirse mediante atributos de tono, brillo (o luminosidad) y colorido (o saturación o croma)”. La óptica, rama de la física que estudia el comportamiento y las propiedades de la luz, sostiene que el color se encuentra en el rango de longitud de onda que va de los 360 nm a los 780 nm aproximadamente, a este rango también se le conoce como “luz visible” (ver figura 2.3).

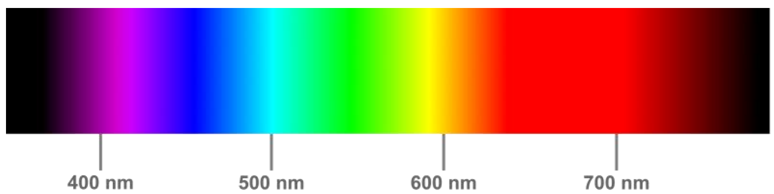

Figura 2.3 Espectro de reflectancia

2.4 Espacios de color

El espacio de color es un sistema para la interpretación y representación del color de una fotografía, imagen, impresión, entre otros. Existen distintos espacios de color, los autores Kahu, Raut y Bhurchandi (2018) los clasifican como: basados en sensores, basados en pantallas, basados en coincidencia de colores, basado en equivalencia de color, basados en modelos de apariencia de colores y sistemas de orden de colores. Dentro de la clasificación de “espacios basados en pantallas” se encuentran los espacios de color aditivo, espacio de color sustractivo y los derivados del espacio de color RGB (ver figura 2.4).

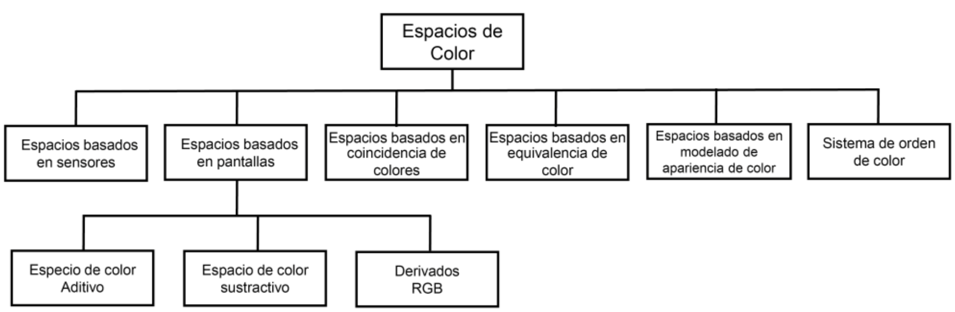

Figura 2.4 Clasificación de los espacios de color según Kahu, Raut y Bhurchandi (2018)

2.5 Sistemas de reproducción de color.

El sistema de color aditivo (ver figura 2.5), basado en los colores primarios rojo, verde y azul (Sheth, 2012), se utiliza generalmente para reproducir color sobre un fondo oscuro, como monitores CRT, pantallas de televisión, etc (Kahu, Raut y Bhurchandi, 2018). Al combinar los tres colores del sistema da como resultado el color blanco.

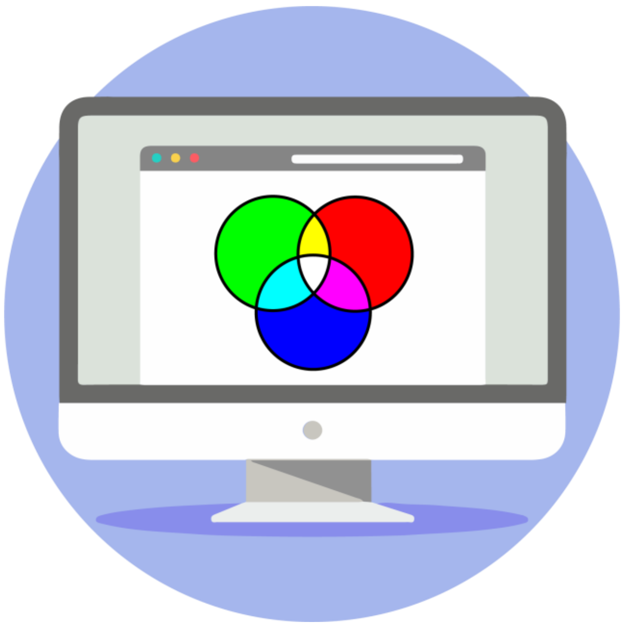

Figura 2.5 *Sistema de color aditivo.*

Los sistemas de color sustractivos combinan un objeto con una fuente de luz la cual se emite hacia el objeto y es reflejada por el mismo; siendo el color resultante el rango de longitud de onda reflejado por el objeto. Este modelo de color sustractivo es utilizado por pintores y por los distintos procesos de impresión que existen; crea color al absorber (restar), ciertas longitudes de onda de color mientras reflejan otras para el espectador (Labrecque, 2020). En la figura 2.6, se observa como la luz va desde de la fuente (el sol), al objeto (la naranja), al ojo y luego al cerebro.

Figura 2.6 *Proceso de percepción de color de modelo sustractivo.*

La industria de la impresión de envases utiliza tradicionalmente colores de proceso CMYK para reproducir imágenes fotográficas y colores directos para reproducir grandes áreas sólidas e información crítica de la marca como logotipos. (Deshpande, 2015); los colores cian, magenta y amarillo con longitudes de onda de corte nítidas se superponen para una mezcla sustractiva (Meyn, 2008), un ejemplo gráfico lo podemos observar en la figura 2.7.

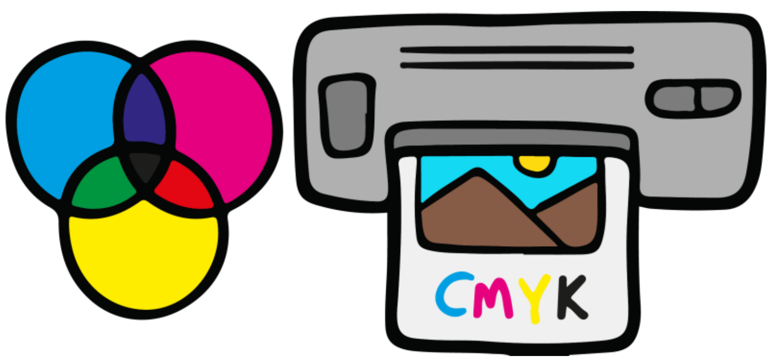

Figura 2.7 *Proceso de impresión con modelo de color sustractivo.*

En el sistema de color CMYK (ver figura 2.8), la misma magnitud de amarillo más cian produce verde, amarillo más magenta produce rojo y tinta cian más magenta produce azul. Se logran varios matices y valores de color anecdóticamente las cantidades relativas de los cuatro colores. Se agrega negro para mejorar la calidad de las imágenes (Kaur y Kaur, 2015).

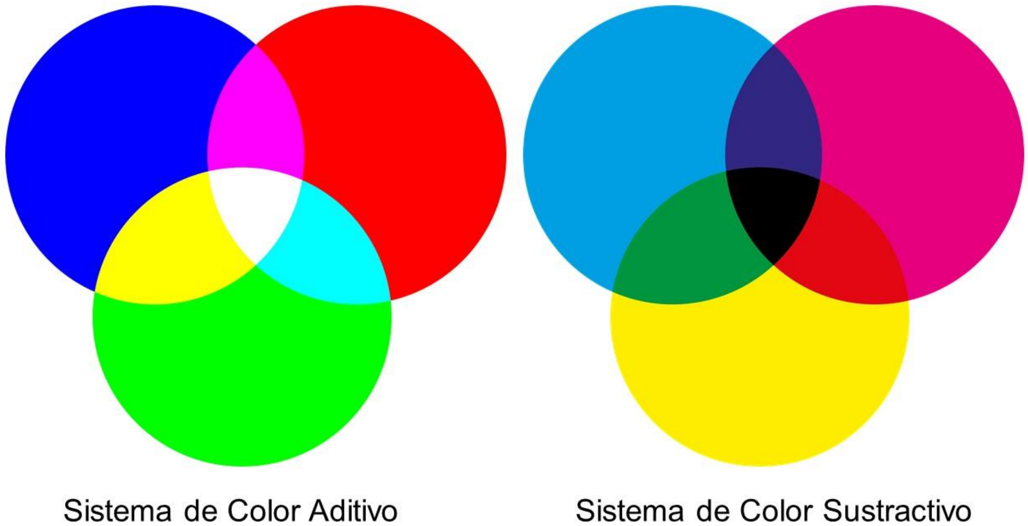

Figura 2.8 *Sistemas de reproducción de color.*

2.6 Atributos de impresión y su medición

Los atributos de impresión son características individuales dentro del proceso de impresión que se pueden y deben monitorear durante su producción para mantener la consistencia del color. Los atributos comúnmente monitoreados son la densidad de la tinta sólida, la ganancia de punto y el contraste de impresión (Dharavath y Hahn, 2009). Para la evaluación y monitoreo de estos atributos se utilizan herramientas de medición como el densitómetro, colorímetro y espectrofotómetro.

2.6.1 Densidad de tinta sólida

Definido por la enciclopedia ColorWiki (2019) “un densitómetro es un dispositivo que mide el grado de oscuridad (la densidad óptica) de un material fotográfico, semitransparente o de una superficie reflectante. Un densitómetro no mide el color, pero mide la densidad”. Asimismo, Verikas y Bacauskiene (2008) mencionan que la densidad, en la industria de las artes gráficas, se considera el principal parámetro adecuado para las evaluaciones de la calidad de impresión durante la ejecución del proceso. Se calcula a partir de la cantidad de luz que se refleja en el sustrato (papel, película, etc.) y la tinta. Es una forma sencilla de evaluar los cambios en el grosor de la película de tinta o la concentración de tinta que se está aplicando. A medida que aumenta el grosor o la concentración de la tinta, se absorbe más luz y se refleja menos luz, por lo que el instrumento informa que la apariencia más oscura es de mayor densidad (X-Rite, 2020); Para Jangra, Saini y Kundu (2014) la densidad generalmente es relacionada con el grosor de la película de tinta, sin embargo, una tinta con una 'carga de pigmento' más alta medirá la misma densidad con un grosor de película de tinta 'más delgada' (ver figura 2.9).

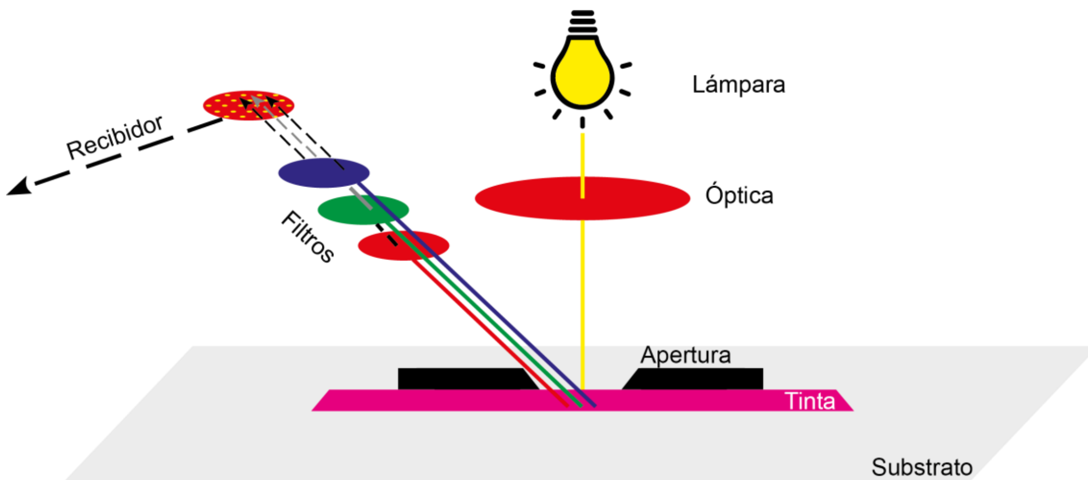

Figura 2.9 *Proceso de medición de un densitómetro*

Debido a la relación que tiene con el grosor de la tinta y la carga pigmentaria, la densidad afecta directamente a la gama de colores. La alta densidad de impresión también podría causar diferentes defectos de impresión (secado y similares) (Tutak, Beytut y Ozcan, 2017). Como lo menciona Sahinbaskan (2002) para obtener un gamut (gama) de color más amplio se debe de obtener una óptima densidad de impresión.

2.6.2 Medición en el espacio de color CIE L*a*b*

Dado que la percepción del color depende principalmente del observador y otras condiciones de iluminación, una evaluación objetiva del color requiere una instrumentación adecuada que siga las recomendaciones instituidas por la CIE (Comisión Internacional de Iluminación, por sus siglas en inglés) (Korifi, *et al.,* 2013). Esto es posible gracias a instrumentos especiales como los espectrofotómetros, y con relación a lo que dice Setchell (2012) "pueden dar resultados confiables en la medición de materiales coloreados cuyas propiedades cambian suavemente a través del espectro".

Para definir objetivamente un color, además de utilizar la instrumentación adecuada, es necesario seguir las recomendaciones establecidas por el CIE. La CIE recomienda una combinación particular de iluminante / observador y espacios de color (CIEXYZ, CIEL * a * b *) con el propósito de estandarizar la definición de color y proporcionar diferencias de color más uniformes en relación con las diferencias visuales (Korifi, *et al.,* 2013). De acuerdo con Izdebska y Thomas (2016) las evaluaciones objetivas del color generalmente se especifican dentro del espacio de color CIEL*a*b*.

El espacio de color CIEL*a*b* (ver figura 2.10), ampliamente utilizado en las industrias de las artes gráficas, implica la mezcla de colores sustractivos. Hay tres parámetros en el modelo CIELAB: L * representa la claridad del color (L * = 0 indica negro, mientras que L * = 100 indica blanco), el eje −a * / + a * representa los colores del verde al rojo, y el eje −b * / + b * representa los colores del azul al amarillo (Mahajan y Bandyopadhyay, 2019). El espacio de color CIEL*a*b* también define el croma (C*) y el tono ($\mathrm{h}_{\mathrm{ab}}^{*}$) en coordenadas polares como:

$$C^{*} = \sqrt{a^{2} + b^{2}} \tag{2.1}$$

$$h_{ab}^{*} = tan^{-1}\left(\frac{b^{*}}{a^{*}}\right) \tag{2.2}$$

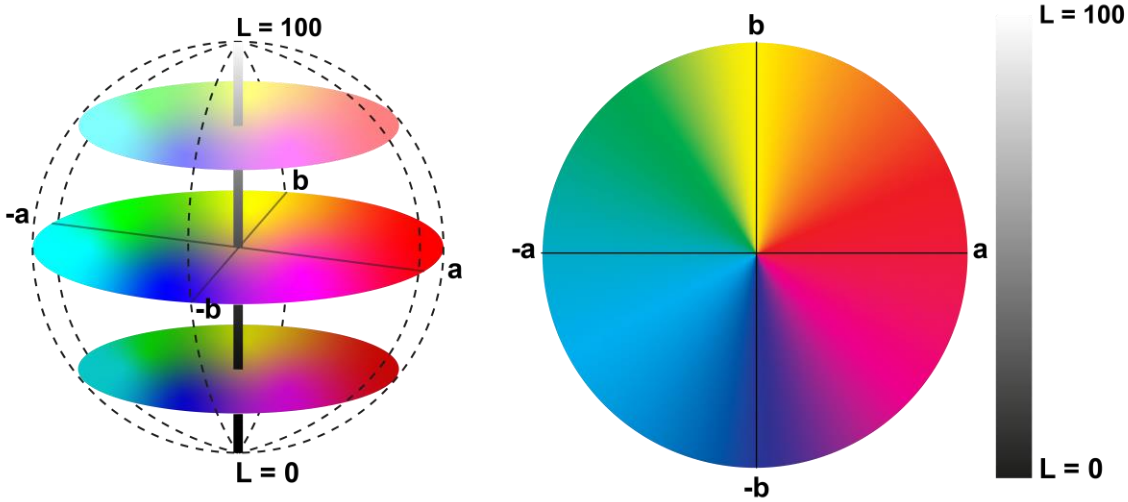

Figura 2.10 *Espacio de color CIEL*a*b* visto en 3D y 2D*

2.7 Gamut de color

El gamut de color está dentro de un espacio de color en materiales como papel, superficies plásticas, madera, metal, cerámicas, entre otros. Se define como una gama de colores alcanzables en un medio de reproducción de color dado (o presente en una imagen en ese medio) bajo un conjunto dado de condiciones de visualización: es un volumen en el espacio de color (Zhao, 2007). Al imprimir en diferentes materiales se crea un gamut de color para cada uno de ellos; en la figura 2.11 se puede apreciar como el gamut forma un volumen dentro del espacio de color CIEL*a*b*, su forma depende de los atributos de impresión previamente descritos en los puntos 2.3.3.1 y 2.3.3.2.

Se entiende que el espacio de color L* a* b*, o CIEL*a*b*, es un estándar internacional para las mediciones de color, adoptado por la Comisión Internacional de Eclairage (CIE) en 1976. L* es el componente de luminancia o luminosidad, que varía de 0 a 100, y los parámetros a* (de verde a rojo) y b* (de azul a amarillo) son los dos componentes cromáticos, que varían de -120 a 120 (Papadakis, *et al.,* 2000; Segnini, et al., 1999; Yam & Papadakis, 2004; León *et al.,* 2006). Es deseable que el *gamut* de color sea lo más grande posible (Camp et al., 2002).

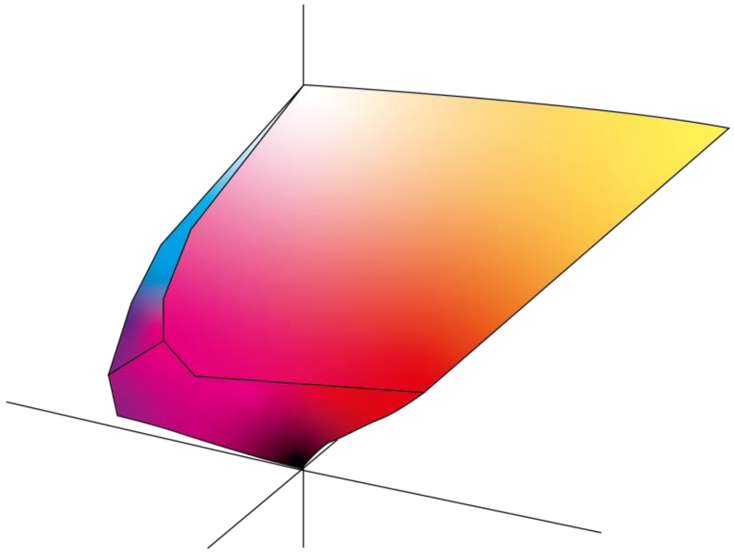

Figura 2.11 *Gamut de color graficado dentro del espacio de color CIE L*a*b**

2.8 Formulación de colores en impresión

La complejidad de una etiqueta inicia cuando como proveedor se tiene que cumplir con un color especial (colores directos), o sea, un tono especificado por el cliente o por PANTONE®, institución que se dedica a la comparación y comunicación del color. Los colores directos son importantes para la identidad de la marca y se utilizan para crear consistencia en muchos productos de marca (Sharma, 2018). Es común que algunas impresiones utilicen colores directos además de las tintas de proceso CMYK. Los colores directos son colores específicos, a menudo indicados de acuerdo con un sistema patentado, como PANTONE, que se formulan e imprimen con una tinta separada. Los colores directos se imprimen con una tinta única específica y, por lo tanto, no tienen una gama limitada en comparación con un color creado a partir de una combinación de sobreimpresiones CMYK (Sharma y Saymour, 2019). Como consecuencia de imprimir colores con tintas específicas se obtiene un incremento en tiempos de configuración de máquina, así como un aumento en los desperdicios.

En los últimos años se ha buscado reducir la práctica de impresión de colores directos con el fin de disminuir los costos de producción. A partir de esto se ha creado el método de gamut extendido. La impresión de gamut extendido es aplicable a todos los procesos de impresión principales: inyección de tinta, litografía offset, huecograbado, flexografía y xerografía. La impresión de gama expandida es un enfoque en la reproducción del color que amplía la gama de colores de los procesos de impresión CMYK convencionales mediante el uso de colorantes adicionales, como las tintas naranja, verde y violeta. (Sharma y Saymour, 2019). Un beneficio importante de un sistema de gama expandida es que una prensa puede usar una paleta fija de siete colores y ejecutar varios trabajos diferentes con diferentes colores directos sin ningún lavado de prensa (Sharma, 2018).

2.9 Análisis multivariado

Los métodos estadísticos multivariados involucran una muestra de tamaño *n* de una población, donde a cada elemento de la muestra se le miden o cuantifican *p* variables de interés. De esta forma se obtiene una matriz de datos con dimensión *n x p* (Gutierrez-Pulido *et al.* 2012). Como se menciona en Avendaño-Prieto *et al*. (2014) después de recolectar, organizar y filtrar las observaciones, se realiza la aplicación de uno o varios procedimientos estadísticos para el análisis de los datos obtenidos. Por otro lado, Kerlinger (2002) menciona que un análisis multivariado “agrupa un conjunto de técnicas estadísticas cuyo objetivo principal es estudiar la interacción y /o correlación entre variables, para generar inferencias y predicciones”.

2.10 Estudios Previos

En estudios previos se ha tratado de investigar el efecto que tienen la secuenciación de las tintas en el gamut de color. Patel (2009) comenta que “los datos obtenidos de múltiples prensas pueden usarse para hacer una prueba ANOVA”. En su estudio encuentra que respetando una secuencia de Cyan, Magenta, Amarillo y Negro se obtiene un mayor gamut de impresión.

Un estudio realizado por O’Hara, Congdon y Gasque (2016) tuvo como resultado que la mejor permutación para la impresión de las siete tintas CMYKOGV es la secuenciación “KYOMGVC”; al imprimir con la pasada secuencia se obtuvo como resultado un gamut de color mucho más grande.

Así mismo también se cuenta con un estudio previo realizado por Gamm y Gerard (2020) donde describen métodos para determinar si los colores de dos muestras de un objeto fabricado provienen de la misma distribución de probabilidad. Durante su estudio describen que la prueba T2 de Hotelling es útil sólo si lo que preocupa es que los vectores medios de Lab son diferentes. El tema es que en caso de que los vectores medios de L*a*b* son diferentes el gamut de color se ve afectado.

3. METODOLOGIA

A lo largo de este capítulo se presenta la metodología que se empleará para llevar a cabo el análisis estadístico multivariado con el fin de comparar dos unidades de impresión y comprobar si sus gamas de color son iguales o diferentes.

3.1 Prueba de Hipótesis para la comparación de dos poblaciones

Casella y Berger (2002) definen a una hipótesis como “una declaración sobre un parámetro de la población. El objetivo de una prueba de hipótesis es decidir, en base a una muestra de la población, cuál de dos hipótesis complementarias es verdadera. Las dos hipótesis complementarias en un problema de prueba de hipótesis son llamadas 'Hipótesis nula' e ‘Hipótesis alternativa' y son denotadas como H_0 y H_1”.

En la tabla 3.1 se ha enlistado, en base a distintas literaturas, una serie de pasos a seguir para realizar una prueba de hipótesis para la comparación de dos poblaciones.

1	Identificación y exposición del problema
2	Elección de los parámetros y tamaño de muestra
3	Selección de la variable respuesta[a]
4	Comparación de dos poblaciones multivariadas
5	Análisis estadístico de los datos
6	Conclusiones y recomendaciones

Tabla 3.1 Pasos por seguir para una prueba de hipótesis.

3.1.1 Identificación y exposición del problema

Existen varias herramientas que permiten identificar las variables que están causando un problema, Neyestani (2017) las llama como las siete herramientas antiguas de control de calidad y comenta que son un conjunto de herramientas que se utilizan para mejorar los procesos. Las 7 herramientas que él menciona fueron introducidas por Ishikawa y son 1) Lista de verificación; 2) Gráficos o análisis de tendencias; 3) histogramas; 4) Gráficos de Pareto; 5) Diagramas de causa y efecto; 6) Diagramas de dispersión; 7) Gráficos de control. Para Dobrusskin (2016) “el diagrama de espina de pescado es una manera excelente de representar un modo fácil y estandarizado de investigar las causas subyacentes".

En la figura 3.1 podemos observar el diagrama de pescado que define todas las variables que intervienen en el proceso de impresión flexográfico de etiquetas autoadheribles y que juntas interfieren en la calidad de impresión.

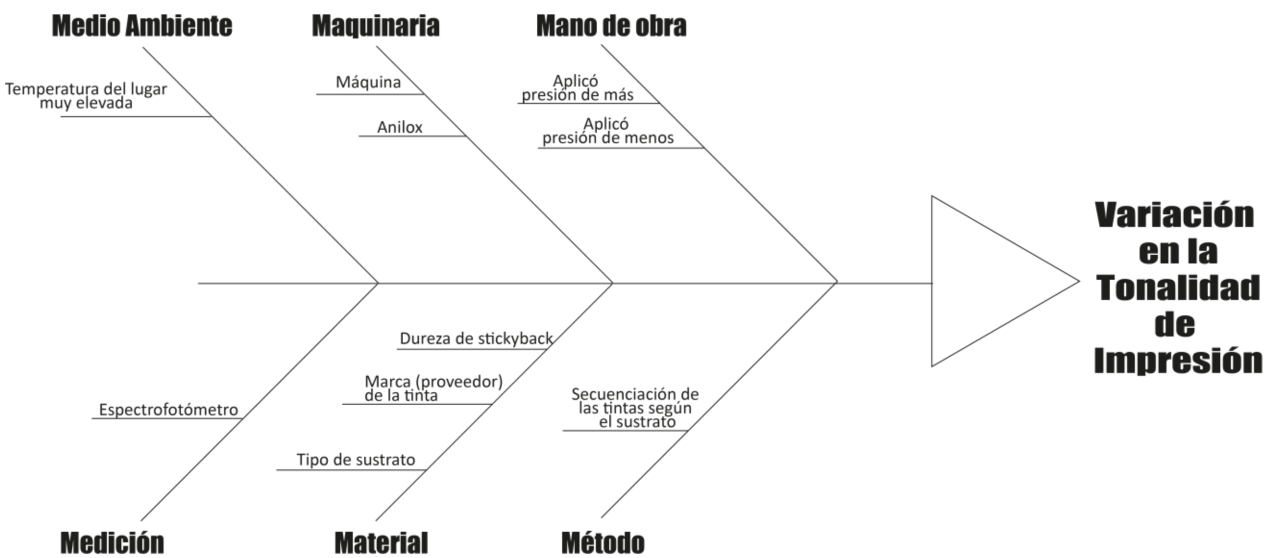

Figura 3.1 *Diagrama de Ishikawa.*

3.1.2 Elección de los parámetros y tamaño de muestra.

En el caso de las tintas Cian, Magenta, Amarillo y Negro se tendrá como parámetro los valores estándar, dentro del espacio de color CIEL*a*b*, definidos por "GRACol®" en el perfil de color "GRACoL2006_Coated1v2". GRACol® es un conjunto de pautas y recomendaciones para ayudar a los compradores de impresión, diseñadores y especificadores a trabajar de manera más eficaz con sus proveedores de impresión" (Idealliance, 2021). Para las tintas naranja, verde y violeta se tendrán como parámetro los valores CIEL*a*b* definidos por "ESKO®". Para el caso de la densidad, el parámetro se estimará por medio de regresión.

Por otro lado, el tamaño de la muestra será de $n = 60$ para cada color.

3.1.3 Selección de la variable respuesta

En el capítulo 2 se describieron los atributos de impresión, los mencionados fueron las coordenadas dentro del espacio CIE L*a*b* y la densidad. Estos atributos son los normalmente utilizados para el control de calidad en la industria de impresión.

3.1.4 Comparación de dos poblaciones multivariadas y análisis estadístico de los datos.

La medida de centralización más utilizada para describir datos multivariantes es el vector de medias, que es un vector de dimensión p cuyos componentes son las medias de cada una de las p variables. Puede calcularse, como el caso escalar, promediando las medidas de cada elemento, que ahora son vectores. (Peña, 2002)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{bmatrix} \overline{x_1} \\ \vdots \\ \overline{x_p} \end{bmatrix} \quad (3.1)$$

Su expresión a partir de la matriz de datos es:

$$\bar{x} = \frac{1}{n}\mathrm{X}'1 \quad (3.2)$$

La relación lineal entre dos variables se mide por la covarianza (Peña, 2002). Se calcula con:

$$s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_{ij})(x_{ij} - \bar{x}_k) \quad (3.3)$$

Y mide su dependencia lineal. Esta información para una variable multivariante puede presentarse de forma compacta en la matriz de varianzas y covarianzas (Peña, 2002). Se define esta matriz como:

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \boldsymbol{\bar{x}})(x_i - \boldsymbol{\bar{x}})' \quad (3.4)$$

Que es una matriz cuadrada y simétrica que contiene en la diagonal las varianzas y fuera de la diagonal las covarianzas entre las variables.

La dependencia lineal entre dos variables se estudia mediante el coeficiente de correlación lineal o simple (Peña, 2002). Este coeficiente para las variables x_j, x_k es:

$$r_{jk} = \frac{s_{jk}}{s_j s_k} \tag{3.5}$$

y tiene las siguientes propiedades:

1. $0 \leq |r_{jk}| \leq 1$
2. Si existe una relación lineal exacta entre las variables, $x_{ij} = a + bx_{ik}$, entonces $|r_{jk}| = 1$
3. r_{jk} es invariante ante transformaciones lineales de las variables.

La dependencia por pares entre las variables se mide por la matriz de correlación. Tiene unos en la diagonal principal y fuera de ella los coeficientes de correlación lineal entre pares de variables, escribiremos:

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \dots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix} \tag{3.6}$$

Se denomina matriz de precisión a la inversa de la matriz de varianzas y covarianzas. Esta matriz juega un papel importante en muchos procedimientos estadísticos. Contiene información sobre la relación multivariante entre cada una de las variables y el resto. (Peña, 2002)

1. Es decir, si llamamos s^{ij} a los elementos de la matriz de precisión:

$$s^{ij} = -\hat{\beta}_{ij}/s_r^2(i) \tag{3.7}$$

Donde $\hat{\beta}_{ij}$ es el coeficiente de regresión de la variable j para explicar la variable i, y $s_r^2(i)$ la varianza residual de la regresión.

2. En la diagonal se tienen las inversas de las varianzas residuales de cada variable en su regresión con el resto. Es decir:

$$s^{ii} = \frac{1}{s_r^2(i)} \tag{3.8}$$

3. Si estandarizamos los elementos de esta matriz para que tenga unos en la diagonal, los elementos fuera de la diagonal, son los coeficientes de correlación parcial entre estas variables. Es decir:

$$r_{ij.R} = -\frac{s^{ij}}{\sqrt{s^{ii}s^{jj}}} \quad (3.9)$$

Donde R se refiere al resto de las variables, es decir el conjunto de $p-2$ variables x_k con $k = 1, \dots, p \; y \; k \neq i, j$.

La T^2 de dos muestras se utiliza para probar la diferencia entre los vectores medios de dos muestras multivariadas (Gamm y Gerard, 2020). Se calcula:

$$T^2 = n \; (\bar{x} - \mu_0) \, S^{-1} \, (\bar{x} - \mu_0) \quad (3.10)$$

Para datos normales multivariados, el estadístico T^2 está relacionado con la distribución F con p y n - p grados de libertad (Gamm y Gerard, 2020):

$$F_{p,n-p} = \frac{n-p}{p\,(n-1)} \, T^2 \quad (3.11)$$

La prueba de razón de verosimilitud está relacionada con la función y el estimador de máxima verosimilitud (Cortez, 2020). Es un método que permite construir regiones de rechazo para contrastar una hipótesis nula simple frente una alternativa compuesta (Olmedo, 2009). Empleándose también para el contraste de igualdad de varias medias. Siendo la $H_0 = \mu_1 = \mu_2 = \cdots = \mu_G = \mu$ y la alternativa $H_1 = no \; todas \; las \; \mu_i \; son \; iguales$ se tiene que:

$\lambda = n \, log \frac{|S|}{|S_w|}$ es el estadístico empleado para el contraste. (3.12)

Donde S está asociada a toda información proporcionada por ambas poblaciones muestrales, y s_w es la matriz de covarianza asociada a una población en particular.

Se rechaza H_0 cuando la diferencia sea grande, es decir, cuando la variabilidad suponiendo que H_0 sea cierta, medida por $|S|$, sea mucho mayor que la variabilidad cuando se permite que las medias de los grupos sean distintas, medida por $|S_w|$

4. IMPLEMENTACIÓN

En este capítulo se presenta la metodología aplicada para llevar a cabo la comparación de impresión de una máquina contra la otra.

4.1 Identificación y exposición del problema

Siendo el color un componente de calidad fundamental para muchos objetos (Gamm y Gerard, 2020), se tiene una necesidad de imprimir etiquetas y empaques donde sus colores sean consistentes, cumplan con un estándar y sobre todo que no presenten variación entre familias de productos o reimpresiones. Es preocupante para el fabricante de etiquetas y empaques las diferencias que las impresiones pudiesen presentar, aun cuando se utiliza el mismo proveedor de tinta para llevar a cabo las impresiones, debido a esto se ha decido realizar un estudio estadístico multivariado para evaluar dos máquinas de impresión flexográfico y determinar mediante la comparación de dos poblaciones multivariadas, que ambas máquinas de impresión presentan los mismos resultados de impresión en la tinta Cyan.

4.2 Realización del estudio estadístico.

Para realizar la comparación de las dos máquinas de impresión se ha impreso la imagen que ilustra la figura 4.1.

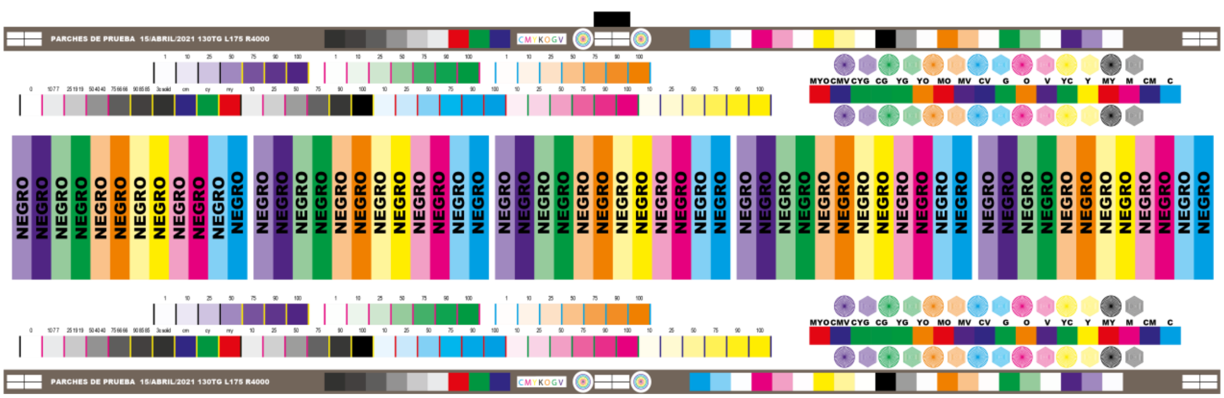

Figura 4.1 Prueba de impresión de las tintas CMYKOGV

El estudio consistió en imprimir 20 hojas por cada máquina y por cada proveedor de tintas (2). La máquina 5 realizó una impresión sobre un material identificado como “polipropileno biorientado”, utilizado comúnmente para envolturas y empaques, es importante aclarar que a dicho material se le imprimió un fondo de tinta blanca puesto que la impresión de la máquina 7 se efectuó en un material identificado como “semigloss” el cual, por su naturaleza, ya presenta un color blanco de origen; éste último se utiliza para la impresión de etiquetas autoadheribles. En total se obtuvieron 2 pares de impresión por cada proveedor de tinta difiriendo únicamente en la máquina de origen. Cabe aclarar que las tintas, aun siendo del mismo proveedor, difieren en su tecnología de secado/curado, esta diferencia entre ambas es que una utiliza lámparas de vapor de mercurio, en lo sucesivo referida como UV estándar (tinta perteneciente a la población 2); y la otra a través de lámparas de curado LED, refereida como lámparas LED (tinta perteneciente a la población 1); los rodillos anilox de ambas máquinas presentan el mismo aporte de tinta.

Las mediciones efectuadas para la evaluación de la impresión se han realizado en las zonas impresas de tinta sólida, cada una de las hojas impresas presentaba un total de 11 zonas de tinta sólida. Se decidió tomar 3 mediciones por cada una de las hojas impresas, teniendo un total de 60 mediciones por cada máquina y por cada color y proveedor de tintas. Las condiciones del espectrofotómetro se muestran en la tabla 1.

INFORMACIÓN DEL ESPECTROFOTÓMETRO	
Condición de medición	M0 (No) - No filter
Iluminante / Observador	D50 / 2°
Estatus de Densidad	ISO Status T
Densidad (referencia)	Absoluta
Espacio de color	CIELab
Equipo	X-Rite eXact

Tabla 4.1 Condiciones de medición

4.3 Discusión y análisis de los resultados

En lo que sigue, se realiza una comparación de tipo multivariado en la reproducción de los colores (cian, magenta, amarillo, negro, naranja, verde y violeta), de dos tipos de máquinas denominadas 5 y 7. La medición consiste de un vector de dimensión cuatro (L, a, b, densidad) en muestras de tamaño 60. El análisis se ejecuta en tres etapas: estimación de parámetros, prueba de hipótesis de una media con respecto a un vector de parámetros dado y prueba de hipótesis del tipo comparación entre vectores de medias. Los cálculos se ejecutan mediante Matlab versión R2020-a en función de sus características de operaciones matriciales.

4.3.1 Proveedor Z

Estimación de parámetros asociados a la máquina 5.

Mediante comandos en Matlab se obtienen los resultados siguientes para el vector (L, a, b, densidad): vector de medias $\bar{x}_{M5}$, matriz de varianzas-covarianzas S_{M5} y matriz de correlaciones R_{M5} muestrales.

Media muestral: $\bar{x}_{M5} = [53.804 \quad -39.392 \quad -47.039 \quad 1.5096]$

Varianza Cian Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 0.2925 & 0.4185 & 0.3405 & -0.0367 \\ 0.4185 & 0.99741 & 0.6419 & -0.0709 \\ 0.3405 & 0.6419 & 0.57874 & -0.0530 \\ -0.0367 & -0.0709 & -0.0530 & 0.0056 \end{bmatrix}$$

Correlación Cian Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & 0.7748 & 0.8275 & -0.9093 \\ 0.7748 & 1 & 0.8449 & -0.9511 \\ 0.8275 & 0.8449 & 1 & -0.9319 \\ -0.9093 & -0.9511 & -0.9319 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión S^{-1} :

Matriz de precisión para la tinta Cian Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 58.958 & 37.264 & 21.67 & 1067.7 \\ 37.264 & 35.715 & 18.716 & 877.11 \\ 21.67 & 18.716 & 23.164 & 600.5 \\ 1067.7 & 877.11 & 600.5 & 24061 \end{bmatrix}$$

Ya que en la diagonal principal de la matriz de precisión se ubican los inversos de las varianzas residuales si la variable i-ésima se explica por las restantes mediante regresión lineal, entonces la variable densidad es la mejor explicada por L a y b al presentarse un inverso de la varianza residual mayor, lo que implica la varianza residual más pequeña que se puede lograr mediante el modelo lineal indicado. Dividiendo el cuarto renglón por 24061 se obtienen los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0444, 0.0365 y 0.0250 respectivamente. Se obtiene la siguiente ecuación:

$$0.0444(\mathrm{L}-\bar{L}) + 0.0365(\mathrm{a}-\bar{a}) + 0.0250(b-\bar{b}) + 1(d-\bar{d}) = 0$$

Por lo que poniendo los valores estimados de los parámetros y despejando la variable densidad se llega a la siguiente ecuación:

$$d = \bar{d} - 0.0444(55.18-\bar{L}) - 0.0365(-37.1-\bar{a}) - 0.0250(-50.06-\bar{b})$$

Cuya varianza de los residuales es $\frac{1}{24061}$ o $4.1561x10^{-5}$ y por lo tanto la desviación estándar de los residuales es 0.00644. Esta ecuación recibe apoyo sólido a partir de los valores de las correlaciones negativas que tiene la variable densidad con respecto a las otras (ver cuarta columna de la matriz de correlación).

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [55.18 \quad -37.1 \quad -50.06]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0444(55.18-\bar{L}) - 0.0365(-37.1-\bar{a}) - 0.0250(-50.06-\bar{b})$$

Mediante los valores que la empresa desea de L, a y b de 55.18, -37.1 y -50.06 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d = 1.5096 - 0.0444(55.18-53.804) - 0.0365(-37.1-(-39.392)) - 0.0250(-50.06-(-47.039))$$

$$d = 1.4404$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (55.18, -37.1, -50.06, 1.4404)$$

$$H_1: \mu \neq (55.18, -37.1, -50.06, 1.4404)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que

$T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 1.1464\text{x}10^4$$

$$F_{4,56} = 2.7203x10^3$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba es casi cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[53.8038, -39.3919, -47.0394, 1.5096]$ se tiene que al estar la densidad por encima de su valor nominal requerido de 1.4404, entonces se espera que, al estar correlacionada negativamente con el resto de las variables, éstas estén por debajo de sus valores nominales, siendo el valor promedio de la variable L igual a 53.8038 sí se cumple esta condición, así mismo el valor promedio para la variable a al ser igual a -39.3919 también se cumple con esta condición, por otro lado la variable b con un valor promedio de -47.0394 no cumple con esta condición (-47.0394 > -50.06). Así la variable b es la problemática en la máquina 5.

Estimación de parámetros asociados a la máquina 7.

Mediante comandos en Matlab se obtienen los resultados siguientes para el vector (L, a, b, densidad): vector de medias $\bar{x}_{M7}$, matriz de varianzas-covarianzas S_{M7} y matriz de correlaciones R_{M7} muestrales.

Media muestral: $\bar{x}_{M7} = [51.2704 \quad -38.2275 \quad -51.7166 \quad 1.7685]$

Varianza Cian Máquina 7, Proveedor Z

$$S_{M7} = \begin{bmatrix} 0.2172 & 0.1384 & 0.0287 & -0.0273 \\ 0.1384 & 0.3324 & -0.0521 & -0.0316 \\ 0.0287 & -0.0521 & 0.1198 & -0.0026 \\ -0.0273 & -0.0316 & -0.0026 & 0.0044 \end{bmatrix}$$

Correlación Cian Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & 0.5152 & 0.1779 & -0.8830 \\ 0.5152 & 1 & -0.2612 & -0.8255 \\ 0.1779 & -0.2612 & 1 & -0.1138 \\ -0.8830 & -0.8255 & -0.1138 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. L se asocia medianamente con a de manera positiva (0.5152) y de manera más fuerte con densidad, pero negativamente (-0.8830). La variable a se relaciona pobremente con b de manera negativa (-0.2612) pero más fuerte con densidad, aunque de manera negativa (-0.8255). La asociación de b con el resto es pobre. La matriz de precisión asociada es:

Precisión Cian Máquina M7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 139.11 & 109.59 & 50.981 & 1678 \\ 109.59 & 102.1 & 49.611 & 1440.5 \\ 50.981 & 49.611 & 32.803 & 691.06 \\ 1678 & 1440.5 & 691.06 & 21363 \end{bmatrix}$$

De nueva cuenta la última fila de la matriz de precisión de la máquina 7 se puede escribir como 21363 * (0.0785, 0.0674, 0.0323, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{21363}$ ó $4.681x10^{-5}$ con una desviación estándar de los residuales igual a 0.0068 y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0785, 0.0674, 0.0323 respectivamente.

Se obtiene la siguiente ecuación:

$$0.0787(\mathrm{L}-\bar{L})+0.0674(\mathrm{a}-\bar{a})+0.0323\left(b-\bar{b}\right)+1\left(d-\bar{d}\right)=0$$

Nuevamente, mediante los valores que la empresa desea de L, a y b de 55.18, -37.1 y -50.06 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ =1.7685-\ 0.0787(55.18-51.2704)-0.0674(-37.1+38.2275)$$
$$-0.0323(-50.06+51.7166)$$

$$d\ =1.3318$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

H_0: $\mu=(55.18,-37.1,-50.06,1.3318)$

H_1: $\mu\neq(55.18,-37.1,-50.06,1.318)$

Mediante $T^2=n\ (\bar{x}-\mu_0)\,S^{-1}\,(\bar{x}-\mu_0)'$ y $F_{p,n-p}=\frac{n-p}{p\,(n-1)}\,T^2$ se obtienen como resultados al aplicar la información obtenida para la máquina 7, lo siguiente:

$$T^2=5.2203\mathrm{x}10^3$$

$$F_{4,56}=\ 1.2387x10^3$$

$$valorprueba=0$$

Se obtiene un resultado similar con respecto a la máquina 5, y al revisar el comportamiento de los valores medios observados se tiene que al estar la media muestral de la densidad por encima del valor deseado (1.7685 > 1.3318) se espera que las demás variables se encuentren por debajo de sus valores deseados, pero no demasiado alejados. La variable L cumple esta condición (51.2704 < 55.18), la variable a también satisface esta condición (-38.22 < -37.1) y b también cumple con la condición (-51.7166 < -50.06) de estar por debajo de los parámetros deseados, pero en conjunto se encuentran alejados del vector deseado**.** La variable b explica de manera pobre a la densidad por lo que puede ser la variable problema (correlación de -0.1138).

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.2549 & 0.2785 & 0.1846 & -0.0320 \\ 0.2785 & 0.6649 & 0.2949 & -0.0513 \\ 0.1846 & 0.2949 & 0.3493 & -0.0278 \\ -0.0320 & -0.0513 & -0.0278 & 0.0050 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\, log \frac{|S|}{|S_w|} = 116.1207$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es aproximadamente cero.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Magenta máquina 5 proveedor Z

Media muestral: $\bar{x}_{M5} = [46.8392 \quad 71.7145 \quad -1.4645 \quad 1.4223]$

Varianza Magenta Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 0.4686 & -0.7332 & 0.0079 & -0.0364 \\ -0.7332 & 1.9503 & 0.1780 & 0.0759 \\ 0.0079 & 0.1780 & 0.3063 & 0.0041 \\ -0.0364 & 0.0759 & 0.0041 & 0.0033 \end{bmatrix}$$

Correlación Magenta Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & -0.7670 & 0.0208 & -0.9266 \\ -0.7670 & 1 & 0.2303 & 0.9464 \\ 0.0208 & 0.2303 & 1 & 0.1275 \\ -0.9266 & 0.9464 & 0.1275 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. L se asocia considerablemente con a de manera negativa (-0.7670) y de manera más fuerte y negativamente con la variable densidad (-0.9266). La variable a se relaciona pobremente con b de manera positiva (0.2303). La asociación de b con el resto es pobre. La matriz de precisión asociada es:

Precisión Magenta Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 85.265 & -42.835 & -2.8488 & 1930.5 \\ -42.835 & 27.085 & -0.13371 & -1096 \\ -2.8488 & -0.13371 & 3.8538 & -33.114 \\ 1930.5 & -1096 & -33.114 & 46875 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 46875 * (0.041184, -0.02338, -0.00070643, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{46875}$ o $2.133x10^{-5}$ con una desviación estándar de los residuales igual a 0.00462 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.041184, -0.02338y -0.00070643 respectivamente. Se obtiene la siguiente ecuación:

$$0.041184(\mathrm{L}-\bar{L})+(-0.02338)(\mathrm{a}-\bar{a})+(-0.00070643)\left(b-\bar{b}\right)+1\left(d-\bar{d}\right)=0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_5 = [48.12 \quad 74.15 \quad -5.05]$$

Se desprende que el mejor valor para la densidad es:

$$d\ =\bar{d}-\ 0.041184(48.12-\bar{L})+0.02338(74.15-\bar{a})+0.00070643\left(-5.05-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 48.12, 74.15 y -5.05 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d\ =1.4223-\ 0.041184(48.12-46.8392)+0.02338(74.15-71.7145)-$$
$$0.0250(-5.05+1.4646)$$

$$d\ =1.424$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (48.12, 74.15, -5.05, 1.424)$$
$$H_1: \mu \neq (48.12, -74.15, -5.0, 1.424)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que

$T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 6672.4$$

$$F_{4,56} = 1583.3$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[46.8392 \quad 71.7145 \quad -1.4645 \quad 1.4223]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.424, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por encima de su valor nominal, siendo el valor promedio de L igual a 46.8392 no se cumple esta condición (46.8392 < 48.12), por otro lado se tiene que al estar correlacionado positivamente con a y con b se espera que éstas estén por debajo de su valor nominal; el valor promedio para la variable a es de 71.7145 por lo tanto si cumple esta condición (71.7145 < 74.15), de manera contraria al tener la variable b un valor promedio de -1.4645 no se cumple esta condición (-1.4645 > -5.05). Por lo tanto, se obtiene que las variables L y b pudiesen ser las de la problemática.

Estimación de parámetros asociados a la tinta Magenta, máquina 7.

Media muestral: $\bar{x}_{M7} = [44.6510 \quad 76.3437 \quad -3.5533 \quad 1.6230]$

Varianza Magenta Máquina 7, Proveedor Z

$$S_{M7} = \begin{bmatrix} 0.0616 & -0.0345 & -0.0857 & -0.0058 \\ -0.0345 & 0.1047 & 0.0907 & 0.0069 \\ -0.0857 & 0.0907 & 0.2393 & 0.0109 \\ -0.0058 & 0.0069 & 0.0109 & 0.0007 \end{bmatrix}$$

Correlación Magenta Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & -0.4293 & -0.7058 & -0.8694 \\ -0.4293 & 1 & 0.5727 & 0.7862 \\ -0.7058 & 0.5727 & 1 & 0.8233 \\ -0.8694 & 0.7862 & 0.8233 & 1 \end{bmatrix}$$

Se observa una asociación considerable entre las variables en función de las correlaciones muestrales obtenidas con la variable densidad. L se asocia considerable y negativamente con un valor igual a -0.8694,por otro lado, las variables a y b se asocian positivamente con la variable densidad, teniendo una correlación con a igual a 0.7862 y con b igual a 0.8233. La matriz de precisión asociada es:

Precisión Magenta Máquina 7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 256.27 & -138.34 & -25.202 & 3727.6 \\ -138.34 & 100.87 & 17.615 & -2318.8 \\ -25.202 & 17.615 & 16.063 & -607.03 \\ 3727.6 & -2318.8 & -607.03 & 62068 \end{bmatrix}$$

De nueva cuenta la última fila de la matriz de precisión de la máquina 7 se puede escribir como 62068 * (0.060057, -0.037359, -0.0097801, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{62068}$ $o\ 1.6111x10^{-5}$ con una desviación estándar de los residuales igual a 0.0040 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son $0.060057, -0.037359, -0.0097801$ respectivamente.

Se obtiene la siguiente ecuación:

$$0.060057(\mathrm{L} - \bar{L}) + (-0.037359)(\mathrm{a} - \bar{a}) + (-0.0097801)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Nuevamente, mediante los valores que la empresa desea de L, a y b de 48.12, 74.15 y -5.05 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d = 1.623 - 0.060057(48.12 - 44.651) + 0.037359(74.15 - 76.344) + 0.0097801(-5.05 + 3.5533)$$

$$d = 1.3181$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$H_0: \mu = (48.12, 74.15, -5.05, 1.3181)$

$H_1: \mu \neq (48.12, 74.15, -5.05, 1.3181)$

Mediante $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ se obtienen como resultados al aplicar la información obtenida para la máquina 7, lo siguiente:

$$T^2 = 19019$$

$$F_{4,56} = 4512.9$$

$$valor\ de\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[44.6510 \quad 76.3437 \quad -3.5533 \quad 1.6230]$ se tiene que, al estar la densidad por encima de su valor nominal requerido de 1.3181, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por debajo de su valor nominal, siendo el valor promedio de L igual a 44.6120 si se cumple esta condición (44.6120 < 48.12), por otro lado se tiene que al estar correlacionado positivamente con a y con b se espera que éstas estén por encima de su valor nominal, el valor promedio para a es igual a 76.3437 por lo tanto si se cumple esta condición (76.3437 > 74.15) y teniendo un valor promedio de -3.5533 para la variable b también cumple con la condición dada (-3.5533 > -5.05).

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log\, \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.2651 & -0.3839 & -0.0389 & -0.0211 \\ -0.3839 & 1.0275 & 0.1343 & 0.0414 \\ -0.0389 & 0.1343 & 0.2728 & 0.0075 \\ -0.0211 & 0.0414 & 0.0075 & 0.0020 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 7$ el estadístico de prueba toma el valor:

$\lambda = n\, log\, \frac{|S|}{|S_w|}$= 322.35

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4.

Bajo estas condiciones el valor de prueba asociado es 0

Por lo que la hipótesis nula se rechaza, significando que ambas máquinas no reproducen las mismas propiedades de impresión.

Estimación de parámetros asociados a la tinta Amarillo, máquina 5.

Media muestral: $\bar{x}_{M5} = [88.0511 \quad -9.2613 \quad 98.0019 \quad 1.1186]$

Varianza Amarillo Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 0.0876 & -0.0082 & -0.2384 & -0.0057 \\ -0.0082 & 0.0234 & 0.2455 & 0.0040 \\ -0.2384 & 0.2455 & 10.0958 & 0.1533 \\ -0.0057 & 0.0040 & 0.1533 & 0.0024 \end{bmatrix}$$

Correlación Amarillo Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & -0.1820 & -0.2534 & -0.3952 \\ -0.1820 & 1 & 0.5046 & 0.5286 \\ -0.2534 & 0.5046 & 1 & 0.9880 \\ -0.3952 & 0.5286 & 0.9880 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La única variable que presenta una correlación altamente positiva con la variable densidad es b con un valor igual a 0.9880. La matriz de precisión asociada es:

Precisión Amarillo Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 332.24 & -116.81 & -181.83 & 12676 \\ -116.81 & 101.36 & 66.084 & -4695.3 \\ -181.83 & 66.084 & 103.74 & -7212.8 \\ 12676 & -4695.3 & -7212.8 & 502160 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 502160 * (0.0252, -0.0094, -0.0144, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{502160}$ ó de $1.9914x10^{-06}$ con una desviación estándar de los residuales igual a 0.0014112 y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0252, -0.0094 y -0.0144 respectivamente. Se obtiene la siguiente ecuación:

$$0.0252(\mathrm{L} - \bar{L}) + (-0.0094)(\mathrm{a} - \bar{a}) + (-0.0144)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [89.29 \quad -4.97 \quad 93.13]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0252(\mathrm{L} - \bar{L}) + 0.0094(\mathrm{a} - \bar{a}) + 0.0144\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 89.29, -4.97 y 93.13 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d = 1.1186 - 0.0252(89.29 - 88.0511) + 0.0094(-4.97 + 9.2613) + 0.0144(93.13 - 98.0019)$$

$$d = 1.0576$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (89.29, -4.97, 93.13, 1.0576)$$

$$H_1: \mu \neq (89.29, -4.97, 93.13, 1.0576)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\ (n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 69137$$

$$F_{4,56} = 16405$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[88.0511 \quad -9.2613 \quad 98.0019 \quad 1.1186]$ se tiene que, al estar la densidad por encima de su valor nominal requerido de 1.0576, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por debajo de su valor nominal, siento el valor promedio de la variable L igual a 88.0511 si se cumple esta condición (88.0511 < 89.29). Por otro lado, se tienen que al estar correlacionada positivamente con a y b se espera que éstas estén por encima de su valor nominal la variable a cuenta con un valor promedio igual a -9.2613 por lo tanto no cumple esta condición (-9.2613 < -4.97) y la variable b teniendo un valor promedio de 98.0019 sí la cumple (98.0019 > 93.13). Por lo tanto, se tiene que la variable a pudiese ser la problemática de la máquina 5.

Estimación de parámetros asociados a la tinta Amarillo, máquina 7.

Media muestral: $\bar{x}_{M7} = [89.2658 \quad -8.2758 \quad 87.7541 \quad 0.9428]$

Varianza Amarillo Máquina 7, Proveedor Z

$$S_{M7} = \begin{bmatrix} 0.0427 & -0.0047 & 0.0984 & 0.0003 \\ -0.0047 & 0.0256 & 0.0743 & 0.0011 \\ 0.0984 & 0.0743 & 2.8784 & 0.0359 \\ 0.0003 & 0.0011 & 0.0359 & 0.0005 \end{bmatrix}$$

Correlación Amarillo Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & -0.1412 & 0.2809 & 0.0632 \\ -0.1412 & 1 & 0.2735 & 0.3281 \\ 0.2809 & 0.2735 & 1 & 0.9751 \\ 0.0632 & 0.3281 & 0.9751 & 1 \end{bmatrix}$$

De nueva cuenta no se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La única variable que presenta una correlación altamente positiva con la variable densidad es b con un valor igual a 0.9751. La matriz de precisión asociada es:

Precisión Amarillo Máquina 7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 1574.1 & -108.13 & -865.02 & 65320 \\ -108.13 & 53.39 & 63.518 & -4911 \\ -865.02 & 63.518 & 482.81 & -36474 \\ 65320 & -4911 & -36474 & 2757900 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 2757900 * (0.0237, -0.0018, -0.0132, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{2757900}$ igual a 3.6259$x10^{-07}$ con una desviación estándar de los residuales igual a 6.0216$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0237, -0.0018 y -0.0132 respectivamente. Se obtiene la siguiente ecuación:

$$0.0237(\mathrm{L}-\bar{L}) + (\,-0.0018)(\mathrm{a}-\bar{a}) + (-0.0132)\left(b-\bar{b}\right) + 1\left(d-\bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = \begin{bmatrix} 89.29 & -4.97 & 93.13 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - 0.0237(\mathrm{L}-\bar{L}) + 0.0018(\mathrm{a}-\bar{a}) + 0.0132\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 89.29, -4.97 y 93.13 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 0.9428 - 0.0237(89.29 - 89.2658) + 0.0018(-4.97 + 8.2758) + 0.0132(93.13 - 87.7541)$$

$$d\ = 1.0191$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (89.29, -4.97, 93.13, 1.0191)$$

$$H_1: \mu \neq (89.29, -4.97, 93.13, 1.0191)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 48679$$

$$F_{4,56} = 11551$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[89.2658 \quad -8.2758 \quad 87.7541 \quad 0.9428]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.0191, entonces se espera que, al estar correlacionadas positivamente las variables L, a y b, éstas deben de estar por debajo de su valor nominal. La variable L cuenta con un valor promedio igual a 89.2658 por lo tanto si se cumple esta condición (89.2658 < 89.29), la variable a cuenta con un valor promedio igual a -8.2658 y así mismo también cumple esta condición (-8.2658 < -4.97) y la variable b obtuvo un valor promedio igual a 87.7541 y de nueva cuenta también cumple se cumple la condición dada (87.7541 < 93.13). Al ser la variable L la que presenta una correlación pobre (0.0632) con respecto a la variable densidad, pudiese ser que esta variable sea la problemática.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.0652 & -0.0065 & -0.0700 & -0.0027 \\ -0.0065 & 0.0245 & 0.1599 & 0.0025 \\ -0.0700 & 0.1599 & 6.4871 & 0.0946 \\ -0.0027 & 0.0025 & 0.0946 & 0.0014 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\, log \frac{|S|}{|S_w|} = 205.34$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0

Por lo que la hipótesis nula se rechaza, significando que ambas máquinas analizadas no reproducen aproximadamente las mismas propiedades de impresión.

Estimación de parámetros asociados a la tinta de color Negro, máquina 5.

Media muestral

$$\bar{x}_{M5} = [17.5774 \quad 0.6839 \quad 3.7355 \quad 1.6089]$$

Varianza Negro Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 1.2655 & -0.1151 & -0.3897 & -0.0474 \\ -0.1151 & 0.0167 & 0.0426 & 0.0043 \\ -0.3897 & 0.0426 & 0.1706 & 0.0146 \\ -0.0474 & 0.0043 & 0.0146 & 0.0018 \end{bmatrix}$$

Correlación Negro Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & -0.7916 & -0.8388 & -0.9998 \\ -0.7916 & 1 & 0.7973 & 0.7897 \\ -0.8388 & 0.7973 & 1 & 0.8361 \\ -0.9998 & 0.7897 & 0.8361 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión S^{-1} :

Precisión Negro Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 1977.5 & 40.752 & 45.381 & 52269 \\ 40.752 & 191.47 & -26.241 & 838.29 \\ 45.381 & -26.241 & 24.393 & 1074.1 \\ 52269 & 838.29 & 1074.1 & 1383800 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 1383800* (0.0378, 0.0006, 0.0008, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{1383800}$ o 7.2265$x10^{-07}$ con una desviación estándar de los residuales igual a 8.5009$x10^{-04}$ y los coeficientes de regresión de las variables

$x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0378, 0.0006 y -0.0008 respectivamente. Se obtiene la siguiente ecuación:

$0.0378(\mathrm{L} - \bar{L}) + 0.0006(\mathrm{a} - \bar{a}) + 0.0008\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = \begin{bmatrix} 15.0100 & 0.2000 & -0.1600 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$d \ = \bar{d} - \ 0.0378(\mathrm{L} - \bar{L}) - 0.0006(\mathrm{a} - \bar{a}) - 0.0008\left(b - \bar{b}\right)$

Mediante los valores que la empresa desea de L, a y b de 15.0100, 0.2000 y -0.1600 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$\begin{aligned} d \ &= 1.6089 - 0.0378(15.0100 - 17.5774) - 0.0006(0.2000 - 0.6839) \\ &\quad - 0.0008(\,-0.1600 - 3.7355) \end{aligned}$$

$d \ = 1.7094$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0 \colon \mu = (15.01, 0.20, -0.16, 1.7094)$$
$$H_1 \colon \mu \neq (15.01, 0.20, -0.16, 1.7094)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que

$T^2 = n \ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 26403$$

$$F_{4,56} = 6265.2$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[17.5774 \quad 0.6839 \quad 3.7355 \quad 1.6089]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.7094, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por encima de su valor nominal, siendo el valor promedio de ésta igual a 17.5774 si se cumple esta condición (17.5774 > 15.01). De lado contrario se tiene que las variables a y b están correlacionadas positivamente con la variable densidad y se espera que sus valores se encuentren por debajo de su valor nominal; la variable a con un valor promedio igual a 0.6839, al contrario que la variable L, no cumple con la condición dada (0.6839 > 0.20), así mismo la variable b tampoco lo cumple ya que su valor promedio es igual a 3.7355 (3.7355 > -0.16). Así, las variables a y b pudiesen ser las de la problemática en la máquina 5.

Estimación de parámetros asociados a la tinta de color Negro, máquina 7.

Media muestral: $\bar{x}_{M7} = [17.3797 \quad 0.6292 \quad 4.1791 \quad 1.6157]$

Varianza Negro Máquina 7, Proveedor Z

$$S_{M7} = \begin{bmatrix} 0.4798 & 0.0007 & -0.1274 & -0.0184 \\ 0.0007 & 0.0096 & -0.0528 & 0 \\ -0.1274 & -0.0528 & 0.4766 & 0.0045 \\ -0.0184 & 0 & 0.0045 & 0.0007 \end{bmatrix}$$

Correlación Negro Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & 0.0106 & -0.2663 & -0.9996 \\ 0.0106 & 1 & -0.7825 & -0.0017 \\ -0.2663 & -0.7825 & 1 & 0.2473 \\ -0.9996 & -0.0017 & 0.2473 & 1 \end{bmatrix}$$

Para este caso no se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La única variable que presenta una correlación altamente negativa con la variable densidad es L con un valor igual a -0.9996. La matriz de precisión asociada es:

Precisión Negro Máquina 7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 10791 & 1542.5 & 402.28 & 2.7903x10^{5} \\ 1542.5 & 520.8 & 92.924 & 39658 \\ 402.28 & 92.924 & 21.40 & 10361 \\ 2.7903x10^{5} & 39658 & 10361 & 7216600 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 7216600 * (0.0387, 0.0055, 0.0014, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{7216600}$ o 1.3857$x10^{-07}$ con una desviación estándar de los residuales igual a 3.7225$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0387, 0.0055 y -0.0014 respectivamente. Se obtiene la siguiente ecuación:

$$0.0387(\mathrm{L}-\bar{L}) + 0.0055(\mathrm{a}-\bar{a}) + (-0.0014)\left(b-\bar{b}\right) + 1\left(d-\bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_7 = \begin{bmatrix} 15.0100 & 0.2000 & -0.1600 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - 0.0387(\mathrm{L}-\bar{L}) - 0.0055(\mathrm{a}-\bar{a}) - 0.0014\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 15.01, 0.20 y -0.16 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 1.6157 - 0.0387(15.01 - 17.3797) - 0.0055(0.20 - 0.6292) - 0.0014(-0.16 - 4.1791)$$

$$d\ = 1.7160$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (15.0100, 0.2000, -0.1600, 1.7160)$$

$$H_1: \mu \neq (15.0100, 0.2000, -0.1600, 1.7160)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n \ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 22803$$

$$F_{4,56} = 5410.9$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[17.3797 \quad 0.6292 \quad 4.1791 \quad 1.6157]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.7160, entonces se espera que, al estar correlacionadas negativamente con las variables L y a, éstas deben de estar por encima de su valor nominal, siendo el valor promedio para la variable L igual a 17.3797 si se cumple esta condición (17.3797 > 15.01), la variable a obtuvo un valor promedio de 0.6292 y también cumple esta condición (0.6292 > 0.20), teniendo en cuenta que aunque cumple con la condición dada, presenta una correlación pobre con valor igual a -0.0017. Por otro lado, se tiene una correlación positiva con la variable b y se espera que ésta esté por debajo de su valor nominal, donde su valor promedio es igual a

4.1791 por lo que no cumple (4.1791 > -0.16). Así la variable b y posiblemente la variable a (por la pobre correlación que presenta) pudiesen ser las de la problemática en la máquina 7.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0 \colon \mu_5 = \mu_7 = \mu$

$H_1 \colon \mu_5 \neq \mu_7 = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log\, \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 1.2655 & -0.1151 & -0.3897 & -0.0474 \\ -0.1151 & 0.0167 & 0.0426 & 0.0043 \\ -0.3897 & 0.0426 & 0.1706 & 0.0146 \\ -0.0474 & 0.0043 & 0.0146 & 0.0018 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\, log\, \frac{|S|}{|S_w|} = 72.588$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Naranja, máquina 5.

Media muestral: $\bar{x}_{M5} = [68.817 \quad 45.341 \quad 75.866 \quad 1.3712]$

Varianza Naranja Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 0.085087 & -0.011858 & -0.093211 & -0.0044616 \\ -0.011858 & 0.14993 & 0.39696 & 0.0078206 \\ -0.093211 & 0.39696 & 1.5812 & 0.032611 \\ -0.0044616 & 0.0078206 & 0.032611 & 0.00075372 \end{bmatrix}$$

Correlación Naranja Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & -0.10499 & -0.25412 & -0.55712 \\ -0.10499 & 1 & 0.81529 & 0.7357 \\ -0.25412 & 0.81529 & 1 & 0.94464 \\ -0.55712 & 0.7357 & 0.94464 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La matriz de precisión asociada es:

Precisión Naranja Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 7159.9 & -12.195 & -4194.4 & 2.2399x10^5 \\ -12.195 & 20.589 & 0.065774 & -288.67 \\ -4194.4 & 0.065774 & 2465.4 & -1.315x10^5 \\ 2.2399x10^5 & -288.67 & -1.315x10^5 & 7.019700x10^6 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 7019700* (0.0319, -0.00004, -0.0187, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{7019700}$ o 1.4246$x10^{-07}$ con una desviación estándar de los residuales igual a 3.7743$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0319, -0.00004 y -0.0187 respectivamente. Se obtiene la siguiente ecuación:

$$0.0319(\mathrm{L} - \bar{L}) + (-0.00004)(\mathrm{a} - \bar{a}) + (-0.0187)(b - \bar{b}) + 1(d - \bar{d}) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_5 = [67 \quad 61.98 \quad 82.26]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0319(\mathrm{L} - \bar{L}) + 0.00004(\mathrm{a} - \bar{a}) + 0.0187(b - \bar{b})$$

Mediante los valores que la empresa desea de L, a y b de 67, 61.98 y 82.26 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d = 1.3712 - 0.0319(67 - 68.8166) + 0.00004(61.98 - 45.3406) + 0.0187(82.26 - 75.8664)$$

$$d = 1.5496$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (67, 61.98, 82.26, 1.5496)$$

$$H_1: \mu \neq (67, 61.98, 82.26, 1.5496)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n \ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 289980$$

$$F_{4,56} = 68809$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[68.817 \quad 45.341 \quad 75.866 \quad 1.3712]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.5496, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por encima de su valor nominal, siendo el valor promedio de L igual a 68.817 si se cumple esta condición (68.817 > 67). De lado contrario, se tiene que las variables a y b están correlacionadas positivamente con la variable densidad y se espera que sus valores se encuentren por debajo de su valor nominal; la variable a obtuvo un valor promedio igual a 45.341, y al igual que la variable L, cumple con la condición dada (45.341 < 61.98), así mismo la variable b cumple ya que, su valor promedio es igual a 75.866 (75.866 < 82.26). Pudiese ser que el problema sean las tres variables debido a su pobre correlación con la variable densidad.

Estimación de parámetros asociados a la tinta Naranja, máquina 7.

Media muestral: $\bar{x}_{M7} = [71.825 \quad 45.287 \quad 62.686 \quad 1.0703]$

Varianza Naranja Máquina 7, Proveedor Z

$$S_7 = \begin{bmatrix} 0.077447 & -0.07707 & -0.14093 & -0.14093 \\ -0.07707 & 0.17735 & 0.31763 & 0.0066706 \\ -0.14093 & 0.31763 & 0.62524 & 0.012864 \\ -0.14093 & 0.0066706 & 0.012864 & 0.00029442 \end{bmatrix}$$

Correlación Naranja Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & -0.6576 & -0.64043 & -0.85113 \\ -0.6576 & 1 & 0.95384 & 0.92311 \\ -0.64043 & 0.95384 & 1 & 0.94814 \\ -0.85113 & 0.92311 & 0.94814 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión:

Precisión Naranja Máquina 7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 13577 & 153.86 & -7941.6 & 5.3092x10^5 \\ 153.86 & 66.766 & -116.94 & 5720.5 \\ -7941.6 & -116.94 & 4672.3 & -3.1112x10^5 \\ 5.3092x10^5 & 5720.5 & -3.1112x10^5 & 2.0796x10^7 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 20796000 * (0.0255, 0.0003, -0.0150, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{20796000}$ o 4.8086$x10^{-08}$ con una desviación estándar de los residuales igual a 2.1929$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0255, 0.0003 y -0.0150 respectivamente. Se obtiene la siguiente ecuación:

$$0.0255(\mathrm{L} - \bar{L}) + 0.0003(\mathrm{a} - \bar{a}) + (-0.0150)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = \begin{bmatrix} 67 & 61.98 & 82.26 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - 0.0255(\mathrm{L} - \bar{L}) - 0.0003(\mathrm{a} - \bar{a}) + 0.0150\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 67, 61.98 y 82.26 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 1.0703 - 0.0255(67 - 71.8248) - 0.0003(61.98 - 45.2866) + 0.0150(82.26 - 62.6860)$$

$$d\ = 1.4817$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (67, 61.98, 82.26, 1.4817)$$

$$H_1: \mu \neq (67, 61.98, 82.26, 1.4817)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 2.1258 \text{ x } 10^5$$

$$F_{4,56} = 5.0444x10^4$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[71.825 \quad 45.287 \quad 62.686 \quad 1.0703]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.4817, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por encima de su valor nominal, siendo el valor promedio de L igual a 71.825 si se cumple esta condición (71.825 > 67). De lado contrario, se tiene que las variables a y b están correlacionadas positivamente con la variable densidad y se espera que sus valores se encuentren por debajo de su valor nominal; el valor promedio para la variable a es igual 45.287, y así como la variable L, si cumple con la condición dada (45.287 < 61.98), así mismo la variable b cumple ya que obtuvo un valor promedio igual a 62.686 (62.686 < 82.26). Pudiese ser que el problema sean las tres variables debido a su pobre correlación con la variable densidad.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0 \colon \mu_{M5} = \mu_{M7} = \mu$

$H_1 \colon \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.0851 & -0.0119 & -0.0932 & -0.0045 \\ -0.0119 & 0.1499 & 0.3970 & 0.0078 \\ -0.0932 & 0.3970 & 1.5812 & 0.0326 \\ -0.0045 & 0.0078 & 0.0326 & 0.0008 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\, log \frac{|S|}{|S_w|} = 166.6143$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es cero. Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Verde, máquina 5.

Media muestral: $\bar{x}_{M5} = [69.641 \quad -56.5 \quad -4.6284 \quad 0.94452]$

Varianza Verde Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 0.31761 & 0.64249 & 0.044999 & -0.017834 \\ 0.64249 & 1.9615 & 0.061083 & -0.045993 \\ 0.044999 & 0.061083 & 0.086439 & -0.0023858 \\ -0.017834 & -0.045993 & -0.0023858 & 0.0011536 \end{bmatrix}$$

Correlación Verde Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & 0.81399 & 0.27158 & -0.93172 \\ 0.81399 & 1 & 0.14834 & -0.96686 \\ 0.27158 & 0.14834 & 1 & -0.23892 \\ -0.93172 & -0.96686 & -0.23892 & 1 \end{bmatrix}$$

No se observa una gran asociación entre todas las variables en función de las correlaciones muestrales obtenidas. Pero sí se observa una alta correlación entre las variables a y $densidad$ con respecto a L. Con respecto a la variable a, L se asocia fuerte y positivamente (0.8139) y con respecto a la variable L es caso contrario ya que, aunque la asociación es más alta ésta es negativa (-0.93172). La matriz de precisión asociada es:

Precisión Verde Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 268.15 & 153.73 & 37.503 & 10352 \\ 153.73 & 96.934 & 25.175 & 6293.3 \\ 37.503 & 25.175 & 19.049 & 1622.9 \\ 10352 & 6293.3 & 1622.9 & 415170 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 415170***** (0.0249, 0.0152, 0.0039, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{415170}$ o 2.4087$x10^{-06}$ con una desviación estándar de los residuales igual a 0.0016 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0249, 0.0152 y 0.0039, respectivamente. Se obtiene la siguiente ecuación:

$$0.0249(\mathrm{L}-\bar{L})+0.0152\ (\mathrm{a}-\bar{a})+0.0039\left(b-\bar{b}\right)+1\left(d-\bar{d}\right)=0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5}=\ [62 \quad -74.98 \quad -1.30]$$

Se desprende que el mejor valor para la densidad es:

$$d\ =\bar{d}-0.0249(\mathrm{L}-\bar{L})-0.0152(\mathrm{a}-\bar{a})-0.0039\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 62, -74.98 y -1.30 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$\begin{aligned} d\ &=0.9445-0.0249(62-69.6406)-0.0152(-74.98+56.5001) \\ &\quad -0.0039(-1.30+4.6284) \end{aligned}$$

$$d\ =1.4022$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu=(62,-74.98,-1.30,1.4022)$$

$$H_1: \mu\neq(62,-74.98,-1.30,1.4022)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que

$T^2=n\ (\bar{x}-\mu_0)\,S^{-1}\,(\bar{x}-\mu_0)'$ y $F_{p,n-p}=\frac{n-p}{p\ (n-1)}\ T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 2.5784\text{x}10^4$$

$$F_{4,56} = 6.1181x10^3$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $\bar{x} = [69.641 \quad -56.5 \quad -4.6284 \quad 0.94452]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.4022, entonces se espera que, al estar correlacionadas negativamente las variable L, a y b, éstas deben de estar por encima de su valor nominal, $\bar{L} = 69.641$ si cumple esta condición (69.641 > 62), la variable $\bar{a} = -56.5$ también cumple esta condición (-56.5 > -74.98) pero la variable $\bar{b} = -4.6284$ no cumple (-4.6284 < - 1.30). Así la variable b es la problemática en la máquina 5 debido a que su valor no se encontraba por encima del nominal.

Estimación de parámetros asociados a la tinta Verde, máquina 7.

Media muestral: $\bar{x}_{M7} = [64.499 \quad -67.308 \quad -4.6513 \quad 1.2777]$

Varianza Verde Máquina 7, Proveedor Z

$$S_{M7} = \begin{bmatrix} 0.10106 & 0.06548 & -0.00096223 & -0.0050186 \\ 0.06548 & 0.3847 & -0.024316 & -0.010805 \\ -0.00096223 & -0.024316 & 0.016531 & 0.00039278 \\ -0.0050186 & -0.010805 & 0.00039278 & 0.00042208 \end{bmatrix}$$

Correlación Verde Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & 0.33209 & -0.023542 & -0.76843 \\ 0.33209 & 1 & -0.30492 & -0.84795 \\ -0.023542 & -0.30492 & 1 & 0.1487 \\ -0.76843 & -0.84795 & 0.1487 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La densidad se se asocia medianamente con a de manera negativa (-0.84795). La matriz de precisión asociada es:

Precisión Verde Máquina 7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 313.87 & 196.27 & 101.16 & 8662.5 \\ 196.27 & 133.21 & 72.494 & 5676.4 \\ 101.16 & 72.494 & 102.61 & 2963.1 \\ 8662.5 & 5676.4 & 2963.1 & 2.4793x10^{5} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 247930 * (0.0349, 0.0229, 0.0120, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{247930}$ o 4.0334$x10^{-06}$ con una desviación estándar de los residuales igual a 0.0020083 y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0349, 0.0229 y 0.0120 respectivamente. Se obtiene la siguiente ecuación:

$$0.0349(\mathrm{L} - \bar{L}) + 0.0229(\mathrm{a} - \bar{a}) + 0.0120\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [62 \quad -74.98 \quad -1.30]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0349(\mathrm{L} - \bar{L}) - 0.0229(\mathrm{a} - \bar{a}) - 0.0120\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 62, -74.98 y -1.30 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d = 1.2777 - 0.0349(62 - 64.499) - 0.0229(-74.98 + 67.308) - 0.0120(-1.30 + 4.6513)$$

$$d = 1.5006$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0 : \mu = (62, -74.98, -1.30, 1.5006)$$

$$H_1 : \mu \neq (62, -74.98, -1.30, 1.5006)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 44237$$

$$F_{4,56} = 10497$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[64.499 \quad -67.308 \quad -4.6513 \quad 1.2777]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.5006, entonces se espera que, al estar correlacionadas negativamente las variables L y a, éstas deben de estar por encima de su valor nominal, al obtener un valor promedio igual a 64.499 la variable L si cumple esta condición (64.499 > 62), así mismo la variable a con un valor promedio igual a -67.308 también cumple esta condición (-67.308 > 74.98). Por otro lado, se tiene que una correlación positiva pero pobre igual a 0.1487 con la variable b y se se espera que ésta esté por debajo de su valor nominal siendo su valor promedio igual a -4.6513 sí cumple (-4.6513 < -1.30) pero pudiese ser que el problema se origina por la pobre correlación que se tiene con la variable densidad.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.3176 & 0.6425 & 0.0450 & -0.0178 \\ 0.6425 & 1.9615 & 0.0611 & -0.0460 \\ 0.0450 & 0.0611 & 0.0864 & -0.0024 \\ -0.0178 & -0.0460 & -0.0024 & 0.0012 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\, log \frac{|S|}{|S_w|} = 66.1347$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Violeta, máquina 5.

Media muestral: $\bar{x}_{M5} = [32 \quad 42.336 \quad -62.116 \quad 1.2632]$

Varianza Violeta Máquina 5, Proveedor Z

$$S_{M5} = \begin{bmatrix} 0.33854 & -0.37632 & 0.31563 & -0.011646 \\ -0.37632 & 0.55462 & -0.49762 & 0.013475 \\ 0.31563 & -0.49762 & 0.46542 & -0.011318 \\ -0.011646 & 0.013475 & -0.011318 & 0.0004039 \end{bmatrix}$$

Correlación Violeta Máquina 5, Proveedor Z

$$R_{M5} = \begin{bmatrix} 1 & -0.86848 & 0.79515 & -0.99596 \\ -0.86848 & 1 & -0.97943 & 0.90034 \\ 0.79515 & -0.97943 & 1 & -0.82552 \\ -0.99596 & 0.90034 & -0.82552 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión:

Precisión Violeta Máquina 5, Proveedor Z

$$S_{M5}^{-1} = \begin{bmatrix} 2696.5 & -1207.1 & -781.94 & 96112 \\ -1207.1 & 666.87 & 452.43 & -44377 \\ -781.94 & 452.43 & 316.37 & -28775 \\ 96112 & -44377 & -28775 & 3.448x10^{6} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 3448000***** (0.0279, -0.0129, -0.0083, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{3448000}$ o 2.9002$x10^{-07}$ con una desviación estándar de los residuales igual a 0.0005 y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0279, -0.0129 y -0.0083 respectivamente. Se obtiene la siguiente ecuación:

$$0.0279(\mathrm{L} - \bar{L}) + (-0.0129)(\mathrm{a} - \bar{a}) + (-0.0083)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [21 \quad 48.74 \quad -64.68]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0279(\mathrm{L} - \bar{L}) + 0.0129(\mathrm{a} - \bar{a}) + 0.0083\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 21, 48.74 y -64.68 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d = 1.2632 - 0.0279(21 - 31.9997) + 0.0129(48.74 - 42.3362) + 0.0083(-64.68 + 62.1162)$$

$$d = 1.6308$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (21, 48.74, -64.68, 1.6308)$$

$$H_1: \mu \neq (21, 48.74, -64.68, 1.6308)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 4.5602\mathrm{x}10^4$$

$$F_{4,56} = 1.0821x10^4$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[32 \quad 42.336 \quad -62.116 \quad 1.2632]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.6308, entonces se espera que, al estar correlacionadas negativamente las variables L y b, éstas deben de estar por encima de su valor nominal, siendo el valor promedio de L igual a 32 si cumple esta condición (32 > 21), la variable b con un valor promedio de -62.116 también cumple esta condición (-62.116 > -64.68). De lo contrario, la variable a esta correlacionada positivamente, por lo que ésta debe de estar por debajo de su valor nominal, siendo 42.336 su valor promedio y al igual que L y b cumple esta condición (42.336 < 48.74).

Estimación de parámetros asociados a la tinta Violeta, máquina 7.

Media muestral: $\bar{x}_{M7} = [37.4300 \quad 43.1924 \quad -55.9244 \quad 1.1718]$

Varianza Violeta Máquina 7, Proveedor Z

$$S_{M7} = \begin{bmatrix} 86.989 & 3.6063 & 296.37 & -0.30292 \\ 3.6063 & 3.2488 & 14.841 & 0.045481 \\ 296.37 & 14.841 & 1018.1 & -0.94177 \\ -0.30292 & 0.045481 & -0.94177 & 0.002434 \end{bmatrix}$$

Correlación Violeta Máquina 7, Proveedor Z

$$R_{M7} = \begin{bmatrix} 1 & 0.21452 & 0.99588 & -0.65832 \\ 0.21452 & 1 & 0.25806 & 0.51146 \\ 0.99588 & 0.25806 & 1 & -0.59826 \\ -0.65832 & 0.51146 & -0.59826 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. L se asocia altamente con b de manera positiva (0.9959). La asociación con el resto de las variables es pobre. La matriz de precisión asociada es:

Precisión Violeta Máquina 7, Proveedor Z

$$S_{M7}^{-1} = \begin{bmatrix} 143.69 & -68.455 & -35.983 & 5238.7 \\ -68.455 & 33.881 & 17.08 & -2543.8 \\ -35.983 & 17.08 & 9.0158 & -1308.9 \\ 5238.7 & -2543.8 & -1308.9 & 1.9346x10^5 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 193460 * (0.0271, -0.0131, -0.0068, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{193460}$ o 5.169$x10^{-06}$ con una desviación estándar de los residuales igual a 0.0022735 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0271, -0.0131 y -0.0068 respectivamente. Se obtiene la siguiente ecuación:

$$0.0271(\mathrm{L} - \bar{L}) + (-0.0131)(\mathrm{a} - \bar{a}) + (-0.0068)(b - \bar{b}) + 1(d - \bar{d}) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [21 \quad 48.74 \quad -64.68]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0271(\mathrm{L} - \bar{L}) + 0.0131(\mathrm{a} - \bar{a}) + 0.0068(b - \bar{b})$$

Mediante los valores que la empresa desea de L, a y b de 21, 48.74 y -64.68 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d = 1.1718 - 0.0271(21 - 37.43) + 0.0131(48.74 - 43.1924) + 0.0068(-64.68 + 55.9244)$$

$$d = 1.6304$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (21, 48.74, -64.68, 1.6304)$$

$$H_1: \mu \neq (21, 48.74, -64.68, 1.6304)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 17885$$

$$F_{4,56} = 4244$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[37.4300 \quad 43.1924 \quad -55.9244 \quad 1.1718]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.6304, entonces se espera que, al estar correlacionadas negativamente las variables L y b, éstas deben de estar por encima de su valor nominal, L con un valor promedio igual a 37.43 si cumple esta condición (37.43 > 21), la variable b con un valor promedio igual a -55.9244 también cumple esta condición (-55.9244 > -64.68). De lo contrario, la variable a esta correlacionada positivamente, por lo que ésta debe de estar por debajo de su valor nominal, siendo el valor promedio de esta variable igual a 43.1934 y al igual que L y b cumple esta condición (43.1934 < 48.74).

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0 : \mu_{M5} = \mu_{M7} = \mu$

$H_1 : \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\ log\frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 43.6636 & 1.6150 & 148.3430 & -0.1573 \\ 1.6150 & 1.9017 & 7.1719 & 0.0295 \\ 148.3430 & 7.1719 & 509.2888 & -0.4765 \\ -0.1573 & 0.0295 & -0.4765 & 0.0014 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\ log\frac{|S|}{|S_w|} = 929.8820$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

4.3.2 Proveedor N

Estimación de parámetros asociados a la tinta Cian, máquina 5.

Media muestral: $\bar{x}_{M5} = [55.524 \quad -41.996 \quad -45.58 \quad 1.4943]$

Varianza Cian Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 0.33588 & 0.25643 & 0.30308 & -0.034229 \\ 0.25643 & 0.83268 & 0.3784 & -0.050871 \\ 0.30308 & 0.3784 & 0.40756 & -0.03863 \\ -0.034229 & -0.050871 & -0.03863 & 0.0045156 \end{bmatrix}$$

Correlación Cian Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & 0.48489 & 0.81917 & -0.87892 \\ 0.48489 & 1 & 0.64956 & -0.82961 \\ 0.81917 & 0.64956 & 1 & -0.90049 \\ -0.87892 & -0.82961 & -0.90049 & 1 \end{bmatrix}$$

Se observa una correlación alta entre las variables b y densidad con respecto a L siendo positiva para b (0.8192) y negativa para densidad (-0.87989). Así mismo, se presenta una fuerte correlación entre b y densidad siendo su valor negativo e igual a -0.9004. La asociación con las demás variables es pobre. La matriz de precisión asociada es:

Precisión Cian Máquina 5, Proveedor N

$$S_{M5}^{-1} = \begin{bmatrix} 140.71 & 78.604 & 39.199 & 2287.5 \\ 78.604 & 48.502 & 25.282 & 1358.5 \\ 39.199 & 25.282 & 26.388 & 807.7 \\ 2287.5 & 1358.5 & 807.7 & 39776 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 39776 * (0.0575, 0.0342, 0.0203, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{39776}$o 2.5141$x10^{-05}$ con una desviación estándar de los residuales igual a 0.0050 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0575, 0.0342 y 0.0203 respectivamente. Se obtiene la siguiente ecuación:

$$0.0575(\mathrm{L}-\bar{L})+0.0342(\mathrm{a}-\bar{a})+0.0203\left(b-\bar{b}\right)+1\left(d-\bar{d}\right)=0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5}=\begin{bmatrix}55.18 & -37.10 & -50.06\end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ =\bar{d}-\ 0.0575(\mathrm{L}-\bar{L})-0.0342(\mathrm{a}-\bar{a})-0.0203\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 55.18, -37.10 y -50.06 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d\ =1.4943-0.0575(55.18-55.5237)-0.0342(-37.10+41.9961)-0.0203(-50.06+45.5801)$$

$$d\ =1.4376$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0\colon \mu=(55.18\quad -37.10\quad -50.06\quad 1.4376)$$
$$H_1\colon \mu\neq(55.18\quad -37.10\quad -50.06\quad 1.4376)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2=n\ (\bar{x}-\mu_0)\,S^{-1}\,(\bar{x}-\mu_0)'$ y $F_{p,n-p}=\frac{n-p}{p\,(n-1)}\,T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2=\ 19745$$

$$F_{4,56} = 4685.2$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba casi cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[55.5237 \quad -41.9961 \quad -45.5801 \quad 1.4943]$ se tiene que al estar la densidad por encima de su valor nominal requerido de 1.4376, entonces se espera que, al estar correlacionada negativamente con el resto de las variables, éstas estén por debajo de sus valores nominales, $\bar{L}$ igual a 55.5237 no cumple esta condición (55.5237 > 55.18), $\bar{a}$ igual a -41.9961 cumple con esta condición (-41.9961 < -37.10), $\bar{b}$ igual a -45.5801 no cumple con esta condición (-45.5801 > -50.06). Así las variables L y b son las de la problemática en la máquina 5 (proveedor N).

Estimación de parámetros asociados a la tinta Cian, máquina 7.

Media muestral: $\bar{x}_{M7} = [54.026 \quad -38.859 \quad -49.32 \quad 1.5627]$

Varianza Cian Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 2.1118 & -4.1888 & 3.6714 & -0.032577 \\ -4.1888 & 9.1628 & -7.1774 & 0.026059 \\ 3.6714 & -7.1774 & 7.5885 & -0.095797 \\ -0.032577 & 0.026059 & -0.095797 & 0.0032454 \end{bmatrix}$$

Correlación Cian Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & -0.95223 & 0.91711 & -0.39351 \\ -0.95223 & 1 & -0.86074 & 0.15112 \\ 0.91711 & -0.86074 & 1 & -0.61043 \\ -0.39351 & 0.15112 & -0.61043 & 1 \end{bmatrix}$$

Se observa una correlación alta entre las variables a y b con respecto a L siendo negativa para a (-0.9522) y positiva para b (0.9171). Así mismo, se presenta una fuerte correlación entre a y b siendo su valor negativo e igual a -0.8607. La asociación con las demás variables es pobre. La matriz de precisión asociada es:

Precisión Cian Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 260.4 & 222.08 & 150.72 & 5279.5 \\ 222.08 & 192.46 & 132.65 & 4599.3 \\ 150.72 & 132.65 & 92.969 & 3192 \\ 5279.5 & 4599.3 & 3192 & 1.1059x10^{5} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 110590 * (0.0477, 0.0416, 0.0289, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{110590}$ o 9.0424$x10^{-06}$ con una desviación estándar de los residuales igual a 0.0030071 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0477, 0.0416 y 0.0289 respectivamente. Se obtiene la siguiente ecuación:

$$0.0477(\mathrm{L}-\bar{L}) + 0.0416(\mathrm{a}-\bar{a}) + 0.0289\left(b-\bar{b}\right) + 1\left(d-\bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = \begin{bmatrix} 55.18 & -37.10 & -50.06 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - \ 0.0477(\mathrm{L}-\bar{L}) - 0.0416(\mathrm{a}-\bar{a}) - 0.0289\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 55.18, -37.10 y -50.06 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 1.5627 - 0.0477(55.18 - \ 54.026) - 0.0416(-37.10 + 38.859) - 0.0289(-50.06 + 49.32)$$

$$d\ = \ 1.4558$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0\colon \mu = (55.18 \quad -37.10 \quad -50.06 \quad 1.4558)$$

$$H_1\colon \mu \neq (55.18 \quad -37.10 \quad -50.06 \quad 1.4558)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\ (n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 1716.3$$

$$F_{4,56} = 407.27$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba casi cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[54.026 \quad -38.859 \quad -49.32 \quad 1.5627]$ se tiene que, al estar la densidad por encima de su valor nominal requerido de 1.4558 entonces se espera que, al estar correlacionada negativamente con las variables L y b, éstas deben de estar por debajo de sus valores nominales, $\bar{L}$ igual a 54.026 si cumple esta condición (54.026 < 55.18), $\bar{b}$ igual a -49.32 no cumple con esta condición (-49.32 < -50.06). Por otro lado, se tiene una correlación positiva con la variable a, por lo que se espera que esté por encima de su valor nominal, $\bar{a}$ igual a -38.859 no cumple esta condición (-49.32 > -50.06). Por lo tanto, se obtiene que las variables a y b pudiesen ser las variables problema.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

H_0: $\mu_{M5} = \mu_{M7} = \mu$

H_1: $\mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\,log\,\frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 1.2238 & -1.9662 & 1.9872 & -0.0334 \\ -1.9662 & 4.9978 & -3.3995 & -0.0124 \\ 1.9872 & -3.3995 & 3.9980 & -0.0672 \\ -0.0334 & -0.0124 & -0.0672 & 0.0039 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\,log\,\frac{|S|}{|S_w|} = 287.0635$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es cero.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Magenta, máquina 5.

Media muestral

$$\bar{x}_{M5} = [50.977 \quad 69.584 \quad -6.9248 \quad 1.2161]$$

Varianza Magenta Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 0.34493 & -0.76945 & -0.018693 & -0.024597 \\ -0.76945 & 2.1131 & -0.11776 & 0.060717 \\ -0.018693 & -0.11776 & 0.18589 & -0.00060233 \\ -0.024597 & 0.060717 & -0.00060233 & 0.001847 \end{bmatrix}$$

Correlación Magenta Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & -0.90127 & -0.073821 & -0.97452 \\ -0.90127 & 1 & -0.18789 & 0.9719 \\ -0.073821 & -0.18789 & 1 & -0.032507 \\ -0.97452 & 0.9719 & -0.032507 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión:

Precisión Magenta Máquina 5, Proveedor N

$$S_{M5}^{-1} = \begin{bmatrix} 255.67 & -97.958 & -14.894 & 6620.4 \\ -97.958 & 52.819 & 13.771 & -3036.5 \\ -14.894 & 13.771 & 10.507 & -647.64 \\ 6620.4 & -3036.5 & -647.64 & 1.8832x10^5 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 188320 * (0.035155, -0.016124, -0.0034391, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{188320}$ o 5.3101$x10^{-06}$ con una desviación estándar de los residuales igual a 0.0023044 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0352, -0.0161 y -0.0034respectivamente. Se obtiene la siguiente ecuación:

$$0.0352(\mathrm{L} - \bar{L}) + (\ -0.0161)(\mathrm{a} - \bar{a}) + (-0.0034)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = \begin{bmatrix} 48.12 & 74.15 & -5.05 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d \ = \bar{d} - \ 0.0352(\mathrm{L} - \bar{L}) + 0.0161(\mathrm{a} - \bar{a}) + 0.0034\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 48.12, 74.15 y -5.05 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d \ = 1.2161 - 0.0352(48.12 - 50.977) + 0.0161(74.15 - 69.584) + 0.0034(-5.05 + 6.9248)$$

$$d \ = 1.3966$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (48.12, 74.15, -5.05, 1.3966)$$

$$H_1: \mu \neq (48.12, -74.15, -5.05, 1.3966)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n \ (\bar{x} - \mu_0) \, S^{-1} \, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p \ (n-1)} \, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 2.4045\mathrm{x}10^4$$

$$F_{4,56} = \ 570.5500$$

$$valor \ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[50.977 \quad 69.584 \quad -6.9248 \quad 1.2161]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.3966, entonces se espera que, al estar correlacionada negativamente con las variables L y b, éstas deben de estar por encima de su valor nominal, $\bar{L}$ igual a 50.977 si cumple esta condición (50.977 > 48.12) pero, la variable $\bar{b}$ igual a -6.9248 no cumple esta condición (-6.9248 < -5.05). Por otro lado, se tiene que al estar correlacionado positivamente con a se espera que ésta esté por debajo de su valor nominal la variable $\bar{a}$ igual a 69.584 sí cumple esta condición (69.584<74.15). Así la variable b pudiese ser la variable problemática en la máquina 5 proveedor N.

Estimación de parámetros asociados a la tinta Magenta, máquina 7.

Media muestral: $\bar{x}_{M7} = [50.642 \quad 71.393 \quad -10.507 \quad 1.2299]$

Varianza Magenta Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 0.098323 & -0.13479 & -0.04496 & -0.0059524 \\ -0.13479 & 0.34139 & 0.07762 & 0.010811 \\ -0.04496 & 0.07762 & 0.085206 & 0.003159 \\ -0.0059524 & 0.010811 & 0.003159 & 0.00040645 \end{bmatrix}$$

Correlación Magenta Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & -0.73568 & -0.49121 & -0.94158 \\ -0.73568 & 1 & 0.45511 & 0.91778 \\ -0.49121 & 0.45511 & 1 & 0.53679 \\ -0.94158 & 0.91778 & 0.53679 & 1 \end{bmatrix}$$

Se observa una correlación alta entre la variable $densidad$ con respecto a L siendo negativa para (-0.94158), así mismo se presenta una correlación considerable y negativa para a (-0.7356). Así mismo, se presenta una fuerte correlación entre a y $densidad$ siendo su valor positivo e igual a 0.91778. La asociación con las demás variables es pobre. La matriz de precisión asociada es:

Precisión Magenta Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 1735.6 & -773.15 & -118.9 & 46906 \\ -773.15 & 363.24 & 54.955 & -21411 \\ -118.9 & 54.955 & 24.843 & -3396.1 \\ 46906 & -21411 & -3396.1 & 1285300 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 1285300 ***** (0.0365, -0.0167, -0.0026, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{1285300}$ o 7.7803$x10^{-07}$ con una desviación estándar de los residuales igual a 8.8206$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0365, -0.0167 y -0.0026 respectivamente. Se obtiene la siguiente ecuación:

$$0.0365(\mathrm{L} - \bar{L}) + (\,-0.0167)(\mathrm{a} - \bar{a}) + (-0.0026)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = \begin{bmatrix} 48.12 & 74.15 & -5.05 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} -\ 0.0365(\mathrm{L} - \bar{L}) + 0.0167(\mathrm{a} - \bar{a}) + 0.0026\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 48.12, 74.15 y -5.05 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 1.2299 - 0.0365(48.12 -\ 50.6424) + 0.0167(74.15 - 71.3928) \\ + 0.0026(-5.05 + 10.5074)$$

$$d\ = 1.3823$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0\colon \mu = (48.12, 74.15, -5.05, 1.3823)$$
$$H_1\colon \mu \neq (48.12, -74.15, -5.05, 1.3823)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 22300$$

$$F_{4,56} = 5291.4$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[50.642 \quad 71.393 \quad -10.507 \quad 1.2299]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.3823, entonces se espera que, al estar correlacionadas negativamente con la variable L, ésta debe de estar por encima de su valor nominal, $\bar{L}$ igual a 50.642 si cumple esta condición (50.643 > 48.12). Por otro lado, se tienen que al estar correlacionada positivamente con a y b se espera que éstas estén por debajo de su valor nominal la variable $\bar{a} = 71.393$ sí cumple esta condición (72.393 < 74.15) y la variable $\bar{b}$ igual a -10.507 también la cumple (-10.507 > -5.05). Se tiene que la correlación de b no es tan alta, por lo que esta variable b pudiese ser la problemática en esta máquina.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.2216 & 0.4521 & -0.0318 & -0.0153 \\ 0.4521 & 1.2272 & -0.0201 & 0.0358 \\ -0.0318 & -0.0201 & 0.1355 & 0.0013 \\ -0.0153 & 0.0358 & 0.0013 & 0.0011 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina$ 7 el estadístico de prueba toma el valor:

$$\lambda = n\, log \frac{|S|}{|S_w|} = 205.33$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0.

Por lo que la hipótesis nula se rechaza, significando que ambas máquinas analizadas no reproducen aproximadamente las mismas propiedades de impresión.

Estimación de parámetros asociados a la tinta Amarillo, máquina 5.

Media muestral

$$\bar{x}_{M5} = [88.145 \quad -8.9736 \quad 99.934 \quad 1.1497]$$

Varianza Amarillo Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 0.080441 & -0.027424 & -0.1185 & -0.0040794 \\ -0.027424 & 0.025218 & 0.21397 & 0.0041704 \\ -0.1185 & 0.21397 & 3.1426 & 0.052869 \\ -0.0040794 & 0.0041704 & 0.052869 & 0.00094741 \end{bmatrix}$$

Correlación Amarillo Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & -0.60889 & -0.23569 & -0.46729 \\ -0.60889 & 1 & 0.76007 & 0.8532 \\ -0.23569 & 0.76007 & 1 & 0.96892 \\ -0.46729 & 0.8532 & 0.96892 & 1 \end{bmatrix}$$

Se observa una correlación alta entre la variable $densidad$ con respecto b y a siendo para ambas positiva 0.9689 y 0.8532 respectivamente. La asociación con las demás variables es pobre. La matriz de precisión asociada es:

Precisión Amarillo Máquina 5, Proveedor N

$$S_{M5}^{-1} = \begin{bmatrix} 1534.7 & -358.71 & -906.05 & 58748 \\ -358.71 & 282.57 & 231.16 & -15688 \\ -906.05 & 231.16 & 542.01 & -35165 \\ 58748 & -15688 & -35165 & 2285400 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 2285400 * (0.0257, -0.0069, -0.0154, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{2285400}$ o $4.3756x10^{-07}$ con una desviación estándar de los residuales igual a $6.6148x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0257, -0.0069, -0.0154 respectivamente. Se obtiene la siguiente ecuación:

$$0.0257(\mathrm{L} - \bar{L}) + (-0.0069)(\mathrm{a} - \bar{a}) + (-0.0154)(b - \bar{b}) + 1(d - \bar{d}) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [\,89.29 \quad -4.97 \quad 93.13]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0257(\mathrm{L} - \bar{L}) + 0.0069(\mathrm{a} - \bar{a}) + 0.0154(b - \bar{b})$$

Mediante los valores que la empresa desea de L, a y b de 89.29, -4.97 y 93.13 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d = 1.1497 - 0.0257(89.29 - 88.1448) + 0.0069(-4.97 + 8.9736) + 0.0154(93.13 - 99.9338)$$

$$d = 1.0431$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (89.29, -4.97, 93.13, 1.0431)$$
$$H_1: \mu \neq (89.29, -4.97, 93.13, 1.0431)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 2.3264\text{x}10^5$$

$$F_{4,56} = 5.5202\ \text{x}10^4$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[88.145 \quad -8.9736 \quad 99.934 \quad 1.1497]$ se tiene que, al estar la densidad por encima de su valor nominal requerido de 1.0431, entonces se espera que, al estar correlacionada negativamente la variable L, ésta debe de estar por debajo de su valor nominal, $\bar{L}$ igual a 88.145 si cumple esta condición (88.145 < 89.29), Por otro lado, se tienen que al estar correlacionada positivamente con a y b se espera que éstas estén por encima de su valor nominal la variable $\bar{a}$ igual a -8.9736 no cumple esta condición (-8.9736 < -4.97) y la variable $\bar{b}$ igual a 99.934 si la cumple (99.934>93.13)). Se tiene que la variable a, al no estar por debajo de su valor nominal, pudiese ser la variable problemática de la máquina 5 con el proveedor N.

Estimación de parámetros asociados a la tinta Amarillo, máquina 7.

Media muestral: $\bar{x}_{M7} = [89.393 \quad -8.8296 \quad 84.293 \quad 0.8763]$

Varianza Amarillo Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 0.034968 & -0.010529 & -0.11481 & -0.0020294 \\ -0.010529 & 0.032181 & -0.25576 & -0.0025649 \\ -0.11481 & -0.25576 & 10.886 & 0.12613 \\ -0.0020294 & -0.0025649 & 0.12613 & 0.0014777 \end{bmatrix}$$

Correlación Amarillo Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & -0.31387 & -0.18608 & -0.28231 \\ -0.31387 & 1 & -0.43211 & -0.37194 \\ -0.18608 & -0.43211 & 1 & 0.9945 \\ -0.28231 & -0.37194 & 0.9945 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Sólo entre las variables densidad y b si se presenta una correlación altamente positiva con un valor de 0.9945. La matriz de precisión asociada es:

Precisión Amarillo Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 373.44 & -90.042 & -210.27 & 18305 \\ -90.042 & 82.759 & 69.512 & -5913.3 \\ -210.27 & 69.512 & 132.56 & -11483 \\ 18305 & -5913.3 & -11483 & 9.9571x10^{5} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 995710 * (0.0184, -0.0059, -0.0115, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{995710}$ o $1.0043x10^{-06}$ con una desviación estándar de los residuales igual a 0.0010 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 de 0.0184, -0.0059, -0.0115 respectivamente. Se obtiene la siguiente ecuación:

$$0.0184(\mathrm{L}-\bar{L})+(-0.0059)(\mathrm{a}-\bar{a})+(-0.0115)\left(b-\bar{b}\right)+1\left(d-\bar{d}\right)=0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [\,89.29 \quad -4.97 \quad 93.13]$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - \ 0.0184(\mathrm{L}-\bar{L})+0.0059(\mathrm{a}-\bar{a})+0.0115\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 89.29, -4.97 y 93.13 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 0.8763-0.0184(89.29-89.3935)+0.0059(-4.97+8.8296) + 0.0115(93.13-84.2930)$$

$$d\ = 1.003$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (89.29, -4.97, 93.13, 1.003)$$
$$H_1: \mu \neq (89.29, -4.97, 93.13, 1.003)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que

$T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\ (n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 4.7623\text{x}10^4$$

$$F_{4,56} = 1.1300\ \text{x}10^4$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[89.393 \quad -8.8296 \quad 84.293 \quad 0.8763]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.003, entonces se espera que, al estar correlacionada negativamente con las variables L y a, éstas deben de estar por encima de su valor nominal, $\bar{L}$ igual a 89.393 si cumple esta condición (89.393 > 89.29), por otro lado, la variable $\bar{a}$ igual a-8.8296 no cumple con la condición dada (-8.8296 < -4.97). Al estar correlacionada positivamente con la variable b se espera que ésta esté por debajo de su valor nominal la variable $\bar{b}$ igual a 84.293 sí cumple (84.293 < 93.13). La variable a pudiese ser la problemática al no cumplir con la condición dada.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

H_0: $\mu_{M5} = \mu_{M7} = \mu$

H_1: $\mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log\, \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.0804 & -0.0274 & -0.1185 & -0.0041 \\ -0.0274 & 0.0252 & 0.2140 & 0.0042 \\ -0.1185 & 0.2140 & 3.1426 & 0.0529 \\ -0.0041 & 0.0042 & 0.0529 & 0.0009 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\ log \frac{|S|}{|S_w|} = 317.8827$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Negro, máquina 5.

Media muestral: $\bar{x}_{M5} = [14.129 \quad 1.7776 \quad 3.1753 \quad 1.7476]$

Varianza Negro Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 2.8767 & -0.41178 & -0.62668 & -0.11657 \\ -0.41178 & 0.072677 & 0.10043 & 0.016647 \\ -0.62668 & 0.10043 & 0.20495 & 0.025264 \\ -0.11657 & 0.016647 & 0.025264 & 0.0047283 \end{bmatrix}$$

Correlación Negro Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & -0.90057 & -0.81616 & -0.99948 \\ -0.90057 & 1 & 0.82291 & 0.89802 \\ -0.81616 & 0.82291 & 1 & 0.81157 \\ -0.99948 & 0.89802 & 0.81157 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión S^{-1} :

Precisión Negro Máquina 5, Proveedor N

$$S_{M5}^{-1} = \begin{bmatrix} 367.54 & 22.541 & 16.505 & 8893.4 \\ 22.541 & 83.469 & -12.464 & 328.42 \\ 16.505 & -12.464 & 17.247 & 358.62 \\ 8893.4 & 328.42 & 358.62 & 2.1639x10^5 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 216390 * (0.0411, 0.0015, 0.0017, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{216390}$ o 4.6213$x10^{-06}$ con una desviación estándar de los residuales igual a 0.0021 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0411, 0.0015 y 0.0017 respectivamente. Se obtiene la siguiente ecuación:

$$0.0411(\mathrm{L} - \bar{L}) + 0.0015(\mathrm{a} - \bar{a}) + 0.0017(b - \bar{b}) + 1(d - \bar{d}) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [\, 15.01 \quad 0.20 \quad -0.16]$$

Se desprende que el mejor valor para la densidad es:

$$d \ = \bar{d} - 0.0411(\mathrm{L} - \bar{L}) - 0.0015(\mathrm{a} - \bar{a}) - 0.0017(b - \bar{b})$$

Mediante los valores que la empresa desea de L, a y b de 15.01, 0.20 y -0.16 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d \ = 1.7476 - 0.0411(15.01 - 14.1289) - 0.0015(0.20 - 1.7776) \\ - 0.0017(-0.16 - 3.1753)$$

$$d \ = 1.7193$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (15.01, 0.20, -0.16, 1.7193)$$

$$H_1: \mu \neq (15.01, 0.20, -0.16, 1.7193)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que

$T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 1.3255\text{x}10^4$$

$$F_{4,56} = 3.1453\ \text{x}10^3$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[14.129 \quad 1.7776 \quad 3.1753 \quad 1.7476]$ se tiene que, al estar la densidad por encima de su valor nominal requerido de 1.7193, entonces se espera que, al estar correlacionada negativamente la variable L, ésta debe de estar por debajo de su valor nominal, $\bar{L}$ igual a 14.129 si cumple esta condición (14.129<15.01). De lado contrario se tiene que las variables a y b están correlacionadas positivamente con la variable densidad y se espera que sus valores se encuentren por encima de su valor nominal; la variable $\bar{a}$ igual a 1.7776, si cumple con la condición dada (1.7776>0.20), así mismo la variable b lo cumple ya que, $\bar{b}$ igual a 3.1753 (3.1753>-0.16).

Estimación de parámetros asociados a la tinta Negro, máquina 7.

Media muestral

$$\bar{x}_{M7} = [23.032 \quad 2.1159 \quad 5.0116 \quad 1.4228]$$

Varianza Negro Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 4.4195 & -0.26792 & -1.7748 & -0.15847 \\ -0.26792 & 0.067394 & -0.41576 & -0.0021897 \\ -1.7748 & -0.41576 & 7.1152 & 0.21115 \\ -0.15847 & -0.0021897 & 0.21115 & 0.010418 \end{bmatrix}$$

Correlación Negro Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & -0.49091 & -0.3165 & -0.73849 \\ -0.49091 & 1 & -0.6004 & -0.082636 \\ -0.3165 & -0.6004 & 1 & 0.77551 \\ -0.73849 & -0.082636 & 0.77551 & 1 \end{bmatrix}$$

Caso contrario esta vez no se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La matriz de precisión asociada es:

Precisión Negro Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 1.7163 & 10.605 & 0.51932 & 17.809 \\ 10.605 & 120.18 & 10.364 & -23.486 \\ 0.51932 & 10.364 & 1.4468 & -19.244 \\ 17.809 & -23.486 & -19.244 & 751.93 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 751.93 ***** (0.0237, -0.0312, -0.0256, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{751.93}$ o 0.0013 con una desviación estándar de los residuales igual a 0.0365 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 0.0237, -0.0312, -0.0256 respectivamente. Se obtiene la siguiente ecuación:

$$0.0237(\mathrm{L} - \bar{L}) + (-0.0312)(\mathrm{a} - \bar{a}) + (-0.0256)(b - \bar{b}) + 1(d - \bar{d}) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [\,15.01 \quad 0.20 \quad -0.16]$$

Se desprende que el mejor valor para la densidad es:

$$d \ = \bar{d} - \ 0.0237(\mathrm{L} - \bar{L}) + 0.0312(\mathrm{a} - \bar{a}) + 0.0256\big(b - \bar{b}\big)$$

Mediante los valores que la empresa desea de L, a y b de 15.01, 0.20 y -0.16 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d \ = 1.4228 - \ 0.0237(15.01 - 23.0324) + 0.0312(0.20 - 2.1159) + 0.0256(-0.16 - 5.0116)$$

$$d \ = 1.4206$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (15.01, 0.20, -0.16, 1.4206)$$

$$H_1: \mu \neq (15.01, 0.20, -0.16, 1.4206)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n \ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 6.9884\mathrm{x}10^4$$

$$F_{4,56} = \ 1.6583\ \mathrm{x}10^4$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[23.032 \quad 2.1159 \quad 5.0116 \quad 1.4228]$ se tiene que, al estar la densidad por encima de su valor nominal requerido de 1.4206, entonces se espera que, al estar correlacionadas negativamente con las variables L y a, éstas deben de estar por debajo de su valor nominal, $\bar{L}$ igual a 23.032 no cumple esta condición (23.032>15.01), la variable $\bar{a}$ igual a 2.1159 tampoco cumple esta condición (2.1159>0.20), teniendo en cuenta que aunque cumple con la condición dada, presenta una correlación pobre igual a -0.0826. Por otro lado, se tiene una correlación positiva con la variable b y se espera que ésta esté por encima de su valor nominal, $\bar{b}$ igual a 5.0116 si cumple (5.0116>-0.16). Así las variables L y a son las que pudiesen ser las de la problemática en la máquina 7 con el proveedor N.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

H_0: $\mu_{M5} = \mu_{M7} = \mu$

H_1: $\mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 2.8767 & -0.4118 & -0.6267 & -0.1166 \\ -0.4118 & 0.0727 & 0.1004 & 0.0166 \\ -0.6267 & 0.1004 & 0.2049 & 0.0253 \\ -0.1166 & 0.0166 & 0.0253 & 0.0047 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\ log\frac{|S|}{|S_w|} = 568.8420$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Naranja, máquina 5.

Media muestral

$$\bar{x}_{M5} = [67.9 \quad 50.956 \quad 79.043 \quad 1.4682]$$

Varianza Naranja Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 0.05851 & -0.058543 & -0.31198 & -0.0082739 \\ -0.058543 & 0.24039 & 0.9944 & 0.021804 \\ -0.31198 & 0.9944 & 4.7072 & 0.10461 \\ -0.0082739 & 0.021804 & 0.10461 & 0.0023728 \end{bmatrix}$$

Correlación Naranja Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & -0.49364 & -0.59446 & -0.70221 \\ -0.49364 & 1 & 0.93481 & 0.91296 \\ -0.59446 & 0.93481 & 1 & 0.98979 \\ -0.70221 & 0.91296 & 0.98979 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Si se observa que la variable a con respecto a b y densidad se asocia alta y positivamente con respecto a b su valor es igual a 0.9348 y con respecto a la densidad su valor es igual a 0.9129. Así mismo la variable b también se correlaciona alta y positivamente con la variable densidad obteniendo un valor igual a 0.9898. Caso contrario, la variable L se relaciona negativa y medianamente con todas las variables. La matriz de precisión asociada es:

Precisión Naranja Máquina 5, Proveedor

$$S_{M5}^{-1} = \begin{bmatrix} 1943.3 & 55.272 & -1091.4 & 54382 \\ 55.272 & 36.625 & -43.193 & 1760.3 \\ -1091.4 & -43.193 & 627.58 & -31076 \\ 54382 & 1760.3 & -31076 & 1.5439x10^{6} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 1543900 * (0.0352, 0.0011, -0.0201, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{1543900}$ o 6.4771$x10^{-07}$ con una desviación estándar de los residuales igual a 8.0480$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0352, 0.0011 y -0.0201, respectivamente. Se obtiene la siguiente ecuación:

$$0.0352(\mathrm{L} - \bar{L}) + 0.0011(\mathrm{a} - \bar{a}) + (-0.0201)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [\, 67 \quad 61.98 \quad 82.26]$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - 0.0352(\mathrm{L} - \bar{L}) - 0.0011(\mathrm{a} - \bar{a}) + 0.0201\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 67, 61.98 y 82.26 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d\ = 1.4682 - 0.0352(67 - 67.9001) - 0.0011(61.98 - 50.9560) + 0.0201(82.26 - 79.0427)$$

$$d\ = 1.5521$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (67, 61.98, 82.26, 1.5521)$$

$$H_1: \mu \neq (67, 61.98, 82.26, 1.5521)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 2.2892\text{x}10^5$$

$$F_{4,56} = 5.4320\ \text{x}10^4$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[67.9 \quad 50.956 \quad 79.043 \quad 1.4682]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.5521, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por encima de su valor nominal, $\bar{L}$ igual a 67.9 si cumple esta condición (67.9>67). De lado contrario, se tiene que las variables a y b están correlacionadas positivamente con la variable densidad y se espera que sus valores se encuentren por debajo de su valor nominal; la variable $\bar{a}$ igual a 50.956, y al igual que la variable L, si cumple con la condición dada (50.956<61.98), así mismo la variable b cumple ya que, $\bar{b}$ igual a 79.043 (79.043<82.26).

Estimación de parámetros asociados a la tinta Naranja, máquina 7.

Media muestral: $\bar{x}_{M7} = [68.437 \quad 51.899 \quad 77.851 \quad 1.4252]$

Varianza Naranja Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 0.077286 & -0.066874 & -0.17364 & -0.0059904 \\ -0.066874 & 0.22617 & 0.44953 & 0.010968 \\ -0.17364 & 0.44953 & 1.7368 & 0.040295 \\ -0.0059904 & 0.010968 & 0.040295 & 0.00099971 \end{bmatrix}$$

Correlación Naranja Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & -0.50581 & -0.47394 & -0.6815 \\ -0.50581 & 1 & 0.71724 & 0.72942 \\ -0.47394 & 0.71724 & 1 & 0.96702 \\ -0.6815 & 0.72942 & 0.96702 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Las únicas variables correlacionadas altamente entre sí son b con respecto a densidad b se asocia altamente con la variable densidad con un valor igual a 0.9670. La matriz de precisión asociada es:

Precisión Naranja Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 3543.1 & 189.4 & -2144.7 & 1.056x10^{5} \\ 189.4 & 19.618 & -115.28 & 5566.3 \\ -2144.7 & -115.28 & 1307.2 & -64276 \\ 1.056x10^{5} & 5566.3 & -64276 & 3.1634x10^{6} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 3163400 ***** (0.0334, 0.0018, -0.0203, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{3163400}$ o 3.1612$x10^{-7}$ con una desviación estándar de los residuales igual a 5.6224$x10^{-4}$ y los coeficientes de regresión de las variables

$x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0334, 0.0018 y -0.0203, respectivamente. Se obtiene la siguiente ecuación:

$$0.0334(\mathrm{L}-\bar{L}) + 0.0018(\mathrm{a}-\bar{a}) + (-0.0203)\left(b-\bar{b}\right) + 1\left(d-\bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [\,67 \quad 61.98 \quad 82.26]$$

Se desprende que el mejor valor para la densidad es:

$$d = \bar{d} - 0.0334(\mathrm{L}-\bar{L}) - 0.0018(\mathrm{a}-\bar{a}) + 0.0203\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 67, 61.98 y 82.26 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d = 1.4252 - 0.0334(67-68.4367) - 0.0018(61.98-51.8992) + 0.0203(82.26-77.8510)$$

$$d = 1.5450$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados. Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (67, 61.98, 82.26, 1.5450)$$

$$H_1: \mu \neq (67, 61.98, 82.26, 1.5450)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x}-\mu_0)\, S^{-1}\, (\bar{x}-\mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 4.5006\mathrm{x}10^4$$

$$F_{4,56} = 1.0679 \text{ x}10^4$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[68.437 \quad 51.899 \quad 77.851 \quad 1.4252]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.5450, entonces se espera que, al estar correlacionada negativamente con la variable L, ésta debe de estar por encima de su valor nominal, $\bar{L}$ igual a 68.437 si cumple esta condición (68.437>67). De lado contrario, se tiene que las variables a y b están correlacionadas positivamente con la variable densidad y se espera que sus valores se encuentren por debajo de su valor nominal; la variable $\bar{a} = 51.899$, y al igual que la variable L, si cumple con la condición dada (51.899<61.98), así mismo la variable b cumple ya que, $\bar{b}$ igual a 77.851 (77.851<82.26).

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\ log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.0585 & -0.0585 & -0.3120 & -0.0083 \\ -0.0585 & 0.2404 & 0.9944 & 0.0218 \\ -0.3120 & 0.9944 & 4.7072 & 0.1046 \\ -0.0083 & 0.0218 & 0.1046 & 0.0024 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\ log \frac{|S|}{|S_w|} = 34.2434$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Verde, máquina 5.

Media muestral: $\bar{x}_{M5} = [64.563 \quad -68.342 \quad -3.6692 \quad 1.2861]$

Varianza Verde Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 0.2499 & 0.47249 & -0.028465 & -0.0186 \\ 0.47249 & 1.3158 & -0.029118 & -0.046581 \\ -0.028465 & -0.029118 & 0.010777 & 0.0012633 \\ -0.0186 & -0.046581 & 0.0012633 & 0.0017128 \end{bmatrix}$$

Correlación Verde Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & 0.82397 & -0.54849 & -0.89905 \\ 0.82397 & 1 & -0.24452 & -0.98121 \\ -0.54849 & -0.24452 & 1 & 0.29403 \\ -0.89905 & -0.98121 & 0.29403 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. L se asocia considerablemente alto con a y de manera positiva (0.8239) y de manera más fuerte con densidad, pero negativamente (-0.89905). La variable a se relaciona altamente con b de manera negativa (-0.98121). La asociación de b con el resto es pobre. La matriz de precisión asociada es:

Precisión Verde Máquina 5, Proveedor N

$$S_{M5}^{-1} = \begin{bmatrix} 85.229 & 47.067 & 102.61 & 2129.9 \\ 47.067 & 47.634 & 45.153 & 1773.3 \\ 102.61 & 45.153 & 231.24 & 2171.8 \\ 2129.9 & 1773.3 & 2171.8 & 70337 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 70337 * (0.0303, 0.0252, 0.0309, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{70337}$ o 1.4217$x10^{-05}$ con una desviación estándar de los residuales igual 0.0038 y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 son 0.0303, 0.0252 y 0.0309, respectivamente. Se obtiene la siguiente ecuación:

$$0.0303(\mathrm{L} - \bar{L}) + 0.0252(\mathrm{a} - \bar{a}) + 0.0309\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = [\,62 \quad -74.98 \quad -1.30]$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - 0.0303(\mathrm{L} - \bar{L}) - 0.0252(\mathrm{a} - \bar{a}) - 0.0309\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 62, -74.98 y -1.30 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$d\ = 1.2861 - 0.0303(62 - 64.5630) - 0.0252(-74.98 + 68.3418) \\ - 0.0309(-1.30 + 3.6692)$$

$$d\ = 1.4578$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (62, -74.98, -1.30, 1.4578)$$

$$H_1: \mu \neq (62, -74.98, -1.30, 1.4578)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 4.8938\text{x}10^4$$

$$F_{4,56} = 1.1612\text{x}10^4$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[64.563 \quad -68.342 \quad -3.6692 \quad 1.2861]]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.4578, entonces se espera que, al estar correlacionadas negativamente las variables L y a, éstas deben de estar por encima de su valor nominal, $\bar{L}$ igual a 64.563 si cumple esta condición (64.563>62), la variable $\bar{a}$ igual a -68.342 también cumple esta condición (-68.3425>-74.98) pero la variable b al tener una correlacionada positivamente con la variable densidad ésta debe de estar por debajo de su valor nominal $\bar{b}$ igual a -3.6692 lo cual si cumple con lo esperado (-3.6692<-1.30). Debido a que la variable densidad está muy pobremente correlacionada con la variable b es esta variable la que pudiera estar ocasionando problemas.

Estimación de parámetros asociados a la tinta Verde, máquina 7.

Media muestral

$$\bar{x}_{M7} = [65.917 \quad -64.487 \quad -4.592 \quad 1.1761]$$

Varianza Verde Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 0.1273 & 0.058938 & -0.0079657 & -0.0049693 \\ 0.058938 & 0.43184 & -0.0055026 & -0.010552 \\ -0.0079657 & -0.0055026 & 0.43184 & 6.1722x10^5 \\ -0.0049693 & -0.010552 & 6.1722x10^5 & 0.00036718 \end{bmatrix}$$

Correlación Verde Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & 0.25138 & -0.13463 & -0.72684 \\ 0.25138 & 1 & -0.050494 & -0.83794 \\ -0.13463 & -0.050494 & 1 & 0.019424 \\ -0.72684 & -0.83794 & 0.019424 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La variable a con respecto a densidad se asocia considerablemente alto y de manera negativa (-0.83794). La matriz de precisión asociada es:

Precisión Verde Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 491.48 & 325.4 & 171.62 & 15973 \\ 325.4 & 223.25 & 114.69 & 10800 \\ 171.62 & 114.69 & 96.451 & 5602.2 \\ 15973 & 10800 & 5602.2 & 5.2831x10^5 \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 528310 ***** (0.0302, 0.0204, 0.0106, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{528310}$ o 1.8928$x10^{-6}$ con una desviación estándar de los residuales igual 0.0014 y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 0.0302, 0.0204 y 0.0106, respectivamente. Se obtiene la siguiente ecuación:

$$0.0302(\mathrm{L}-\bar{L})+0.0204(\mathrm{a}-\bar{a})+0.0106\left(b-\bar{b}\right)+1\left(d-\bar{d}\right)=0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [\,62 \quad -74.98 \quad -1.30]$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} -\ 0.0302(\mathrm{L}-\bar{L})-0.0204(\mathrm{a}-\bar{a})-0.0106\left(b-\bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 62, -74.98 y -1.30 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d\ = 1.1761-0.0302(62-65.9173)-0.0204(-74.98+64.4869)$$
$$-0.0106(-1.30+4.5920)$$

$$d\ = 1.4736$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (62, -74.98, -1.30, 1.4736)$$
$$H_1: \mu \neq (62, -74.98, -1.30, 1.4736)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x}-\mu_0)\, S^{-1}\, (\bar{x}-\mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 3.8516\mathrm{x}10^4$$

$$F_{4,56} =\ 9.1395\ \mathrm{x}10^3$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[65.917 \quad -64.487 \quad -4.592 \quad 1.1761]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.4736, entonces se espera que, al estar correlacionadas negativamente las variables L y a, éstas deben de estar por encima de su valor nominal, $\bar{L}$ igual a 65.917 si cumple esta condición (65.917>62), la variable $\bar{a}$ igual a -64.3487 también cumple esta condición (-64.3387> -74.98). Por otro lado, se tiene que una correlación positiva pero pobre (0.019424) con la variable b y se espera que ésta esté por debajo de su valor nominal siendo $\bar{b}$ igual a -4.592 sí cumple (-4.592<-1.30) pero pudiese ser que el problema se origina por la correlación pobre que se tiene con la variable densidad.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

H_0: $\mu_{M5} = \mu_{M7} = \mu$

H_1: $\mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\ log\frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.2499 & 0.4725 & -0.0285 & -0.0186 \\ 0.4725 & 1.3158 & -0.0291 & -0.0466 \\ -0.0285 & -0.0291 & 0.0108 & 0.0013 \\ -0.0186 & -0.0466 & 0.0013 & 0.0017 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\ log \frac{|S|}{|S_w|} = 48.6831$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es $6.7987x10^{-10}$.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

Estimación de parámetros asociados a la tinta Violeta, máquina 5.

Media muestral: $\bar{x}_{M5} = [26.752 \quad 50.324 \quad -58.972 \quad 1.6121]$

Varianza Violeta Máquina 5, Proveedor N

$$S_{M5} = \begin{bmatrix} 0.34319 & -0.20844 & 0.16055 & -0.015893 \\ -0.20844 & 0.32121 & -0.26567 & 0.01209 \\ 0.16055 & -0.26567 & 0.23522 & -0.009342 \\ -0.015893 & 0.01209 & -0.009342 & 0.00076974 \end{bmatrix}$$

Correlación Violeta Máquina 5, Proveedor N

$$R_{M5} = \begin{bmatrix} 1 & -0.62781 & 0.56506 & -0.97782 \\ -0.62781 & 1 & -0.96653 & 0.76891 \\ 0.56506 & -0.96653 & 1 & -0.69427 \\ -0.97782 & 0.76891 & -0.69427 & 1 \end{bmatrix}$$

No se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. La variable L se asocia alta y negativamente con la variable densidad (-0.97782), así mismo la variable a se correlaciona de igual manera con la variable b (-0.96653). La matriz de precisión asociada es:

Precisión Violeta Máquina 5, Proveedor N

$$S_{M5}^{-1} = \begin{bmatrix} 2617.6 & -1657.2 & -923.02 & 68873 \\ -1657.2 & 1115 & 648.58 & -43857 \\ -923.02 & 648.58 & 396.38 & -24434 \\ 68873 & -43857 & -24434 & 1.8156x10^{6} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 1815600 * (0.0379, -0.0242, -0.0135, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{1815600}$ o 5.5078$x10^{-07}$ con una desviación estándar de los residuales igual a 7.4215$x10^{-04}$ y los coeficientes de regresión de las variables $x_1, x_2, y\ x_3$ en una regresión para explicar x_4 son 0.0379, -0.0242 y -0.0135, respectivamente. Se obtiene la siguiente ecuación:

$$0.0379(\mathrm{L} - \bar{L}) + (-0.0242)(\mathrm{a} - \bar{a}) + (-0.0135)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M5} = \begin{bmatrix} 21 & 48.74 & -64.68 \end{bmatrix}$$

Se desprende que el mejor valor para la densidad es:

$$d\ = \bar{d} - 0.0379(\mathrm{L} - \bar{L}) + 0.0242(\mathrm{a} - \bar{a}) + 0.0135\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 21, 48.74 y -64.68 respectivamente, se desprende que el mejor valor de densidad para la máquina 5 es:

$$\begin{aligned} d\ &= 1.6121 - 0.0379(21 - 26.7521) + 0.0242(48.74 - 50.3237) \\ &\quad + 0.0135(-64.68 + 58.9722) \end{aligned}$$

$$d\ = 1.7153$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0: \mu = (21, 48.74, -50.3237, 1.7153)$$
$$H_1: \mu \neq (21, 48.74, -50.3237, 1.7153)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\ (n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 2.3593\text{x}10^5$$

$$F_{4,56} = 5.5984\text{x}10^4$$

$$valor\ prueba = 0$$

Por lo tanto, dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[26.752 \quad 50.324 \quad -58.972 \quad 1.6121]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.7153, se espera que, al estar correlacionadas negativamente las variables L y b, éstas deben de estar por encima de su valor nominal, $\bar{L}$ igual a 26.752 si cumple esta condición (26.752>21) así mismo, la variable b se debiera de encontrar por encima, siendo $\bar{b}$ igual a -58.972 sí cumple (-58.972>-64.68); por otro lado la variable a se correlaciona positivamente con la variable densidad se debiera de encontrar por debajo de su valor estándar, siendo $\bar{a}$ igual a 50.324 no cumple con la condición dada (50.324>48.74) se concluye que la variable a es la del problema.

Estimación de parámetros asociados a la tinta Violeta, máquina 7.

Media muestral: $\bar{x}_{M7} = [30.72 \quad 49.474 \quad -59.304 \quad 1.4433]$

Varianza Violeta Máquina 7, Proveedor N

$$S_{M7} = \begin{bmatrix} 0.26213 & -0.36781 & 0.29688 & -0.012416 \\ -0.36781 & 0.591 & -0.48749 & 0.018069 \\ 0.29688 & -0.48749 & 0.42997 & -0.01449 \\ -0.012416 & 0.018069 & -0.01449 & 0.00059533 \end{bmatrix}$$

Correlación Violeta Máquina 7, Proveedor N

$$R_{M7} = \begin{bmatrix} 1 & -0.93449 & 0.88429 & -0.99393 \\ -0.93449 & 1 & -0.96706 & 0.96329 \\ 0.88429 & -0.96706 & 1 & -0.90566 \\ -0.99393 & 0.96329 & -0.90566 & 1 \end{bmatrix}$$

Se observa una gran asociación entre las variables en función de las correlaciones muestrales obtenidas. Por lo tanto, se espera una relación lineal fuerte. Así, es conveniente obtener la relación lineal que mejor ajuste la asociación, lo cual se logra con la matriz de precisión:

Precisión Violeta Máquina 7, Proveedor N

$$S_{M7}^{-1} = \begin{bmatrix} 3412.8 & -1492.2 & -685.78 & 99776 \\ -1492.2 & 728.54 & 346.81 & -44793 \\ -685.78 & 346.81 & 179.73 & -20455 \\ 99776 & -44793 & -20455 & 2.9443x10^{6} \end{bmatrix}$$

La última fila de la matriz de precisión se puede escribir como 2944300 ***** (0.0339, -0.0152, -0.0069, 1) que indica que la varianza residual de una regresión entre la cuarta variable y las otras 3 es igual a $\frac{1}{2944300}$ o 3.396410^{-07} con una desviación estándar de los residuales de 5.827910^{-04} y los coeficientes de regresión de las variables $x_1, x_2, y\, x_3$ en una regresión para explicar x_4 0.0339, -0.0152 y -0.0069, respectivamente. Se obtiene la siguiente ecuación:

$$0.0339(\mathrm{L} - \bar{L}) + (-0.0152)(\mathrm{a} - \bar{a}) + (-0.0069)\left(b - \bar{b}\right) + 1\left(d - \bar{d}\right) = 0$$

Mediante los valores que la empresa desea para L a y b:

$$\bar{\mu}_{M7} = [\,21 \quad 48.74 \quad -64.68]$$

Se desprende que el mejor valor para la densidad es:

$$d \ = \bar{d} - \ 0.0339(\mathrm{L} - \bar{L}) + 0.0152(\mathrm{a} - \bar{a}) + 0.0069\left(b - \bar{b}\right)$$

Mediante los valores que la empresa desea de L, a y b de 21, 48.74 y -64.68 respectivamente, se desprende que el mejor valor de densidad para la máquina 7 es:

$$d \ = 1.4433 - 0.0339(21 - 30.7195) + 0.0152(48.74 - 49.4735) \\ + 0.0069(-64.68 + 59.3037)$$

$$d \ = 1.7242$$

Prueba de hipótesis del vector media con respecto a los parámetros deseados.

Se establece como hipótesis de trabajo la siguiente:

$$H_0\colon \mu = (21, 48.74, -64.68, 1.7242)$$

$$H_1\colon \mu \neq (21, 48.74, -64.68, 1.7242)$$

Debido a que no conoce la matriz de varianzas-covarianzas poblacional V sino su estimación S, entonces se aplica la variable T^2 de Hotelling con grados de libertad (p, n-1) = (4, 59) que representan el número de variables y tamaño de muestra menos uno. Esta variable se relaciona con la distribución $F_{p,n-p}$ lo que facilita el uso de tablas debido a que generalmente no se tabula la T^2. Se tiene que $T^2 = n\ (\bar{x} - \mu_0)\, S^{-1}\, (\bar{x} - \mu_0)'$ y $F_{p,n-p} = \frac{n-p}{p\,(n-1)}\, T^2$ son las ecuaciones que deben aplicarse. Debe observarse que se aplican operaciones matriciales en la obtención de T^2 y que se obtiene un valor numérico real. Con los datos del problema se tiene:

$$T^2 = 3.3080\mathrm{x}10^5$$

$$F_{4,56} = \ 7.8495\ \mathrm{x}10^4$$

$$valor\ prueba = 0$$

Dado el valor de prueba igual a cero, entonces se rechaza la hipótesis nula, lo que implica que por lo menos uno de los elementos del vector (L, a, b, densidad) se encuentra lejano de su contraparte paramétrico. Dado el vector de medias muestral observado $[30.72 \quad 49.474 \quad -59.304 \quad 1.4433]$ se tiene que, al estar la densidad por debajo de su valor nominal requerido de 1.7242, se espera que, al estar correlacionadas negativamente las variables L y b, éstas deben de estar por encima de su valor nominal, $\bar{L}$ igual a 30.72 si cumple esta condición (30.72>21) así mismo, la variable $\bar{b}$ igual a -59.304 también cumple esta condición (-59.304>-64.68). Al ser la variable a, caso contrario de las variables L y b, ya que presenta una correlación positiva se espera que esté por debajo de su valor nominal, $\bar{a}$ igual a 49.474 no cumple (-49.474>48.74). En conclusión, la variable a es la de la problemática.

Comparación de medias poblacionales entre las máquinas 5 y 7.

Aunque de los resultados anteriores se desprende que difícilmente las dos máquinas presenten patrones similares, se realizará una prueba de comparación de vectores media. La hipótesis asociada es la siguiente:

$H_0: \mu_{M5} = \mu_{M7} = \mu$

$H_1: \mu_{M5} \neq \mu_{M7} = \mu$

El estadístico de prueba, basado en el Método de Razón de Verosimilitudes, es:

$\lambda = n\, log \frac{|S|}{|S_w|}$ donde S es la matriz de varianzas asociada a toda la información proporcionada por ambas máquinas, y S_w es la matriz de covarianzas asociada a una máquina en particular (en este caso sólo se tienen dos poblaciones). De la información obtenida del problema se tiene:

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$ por lo que al aplicar los datos se llega a:

$$S = \begin{bmatrix} 0.3027 & -0.2881 & 0.2287 & -0.0142 \\ -0.2881 & 0.4561 & -0.3766 & 0.0151 \\ 0.2287 & -0.3766 & 0.3326 & -0.0119 \\ -0.0142 & 0.0151 & -0.0119 & 0.0007 \end{bmatrix}$$

Eligiendo $S_w = varianza\ máquina\ 5$ el estadístico de prueba toma el valor:

$$\lambda = n\, log \frac{|S|}{|S_w|} = 139.4198$$

El estadístico sigue aproximadamente una distribución $\chi^2_{p(G-1)}$ donde p es la dimensión del vector analizado (4) y G es el número de poblaciones a comparar (2). Por lo tanto, la distribución corresponde a una χ^2_4. Bajo estas condiciones el valor de prueba asociado es 0.

Por lo que se rechaza la hipótesis nula, lo que implica que las máquinas analizadas no producen las mismas propiedades revisadas.

5. CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS

En este capítulo se presentan las conclusiones obtenidas con el análisis estadístico para la comparación de dos poblaciones multivariadas, así como también se proporcionan algunas recomendaciones para la empresa y los trabajos futuros para darle continuidad al proyecto.

5.1 Conclusiones

En el presente trabajo se realizó un estudio de análisis estadístico multivariado para comparar los atributos medios de impresión de las tintas utilizadas en el área de prensa. Se decidió llevar a cabo un análisis de este tipo porqué, además de involucrar más de una variable dependiente y más de una variable independiente, también observa y encuentra, patrones y correlaciones entre varias variables simultáneamente.

El objetivo de este estudio fue el comparar ambas máquinas de impresión y concluir en qué tintas se estaba presentando una diferencia significativa. Para ello, se realizó una prueba donde se imprimió con un set de tintas compuesto por “Cian, Magenta, Amarillo, Negro, Naranja, Verde y Violeta” en cada una de las máquinas

Con los resultados obtenidos se encontró que, al evaluar, uno a uno en proveedor, cada tinta, los atributos de impresión no son aproximadamente igual para ninguno de los colores impresos, teniendo como resultado dos gamas (gamuts) de color diferentes para cada máquina.

Al contar con valores nominales que la empresa desea, también se realizó una comparación estadística para cada máquina y proveedor, obteniendo como resultado que los estadísticos no resultaron ser aproximadamente igual.

Así mismo, al realizar los cálculos de correlación se encontró que para la mayoría de los casos la variable “L” o “luminosidad” presentaba una correlación con la variable “densidad”, en algunos donde los colores visualmente son tonos “claros” esta correlación se presentó negativa y para los casos “oscuros” está correlación se presentó positiva. Gracias a que se observó una gran asociación entre las variables en función de las correlaciones muestrales fue posible obtener una ecuación para determinar el valor de “densidad” nominal que la empresa necesita en función a los valores deseados.

5.2 Recomendaciones y trabajos futuros

Al haberse realizado un análisis estadístico de comparación de dos poblaciones solamente se encontró que los colores obtenidos en impresión no presentan vectores medios iguales para ambas máquinas y efectivamente esto conlleva a que haya variaciones tonales entre familias de productos lo cual es una “no conformidad” que la empresa busca eliminar.

Todas las tintas evaluadas en este estudio son formuladas por el proveedor, esto conlleva a que en caso de que la empresa se quiera ajustar a los parámetros de impresión deseados es necesario retroalimentar al proveedor y exigirle que se apegue a dichos parámetros.

Una recomendación adicional sería que la empresa llevara a cabo un diseño experimental multivariado, puesto que así se identificarían la combinación de los factores que afectan directamente en la gama (gamut) de color de cada una de las máquinas.

6. REFERENCIAS

Ampuero, O., & Vila, N. (2006). Consumer perceptions of product packaging. Journal of Consumer Marketing, 23(2), 100–112. doi:10.1108/07363760610655032

Avendaño Prieto, B., Avendaño, G., & Cruz, W. y Cárdenas-Avendaño, A. (2014). Guía de referencia para investigadores no expertos en el uso de estadística multivariada. Diversitas. 10. 13. 10.15332/s1794-9998.2014.0001.01.

Casella, George, and Roger L. Berger. 2002. Statistical Inference. 2nd ed. Pacific Grove, CA: Thompson Learning.

Colorwiki, 2019. Disponible en: http://www.colorwiki.com/wiki/Densitometer

Creusen, M. E. H., & Schoormans, J. P. L. (2005). The Different Roles of Product Appearance in Consumer Choice*. Journal of Product Innovation Management, 22(1), 63–81. doi:10.1111/j.0737-6782.2005.00103.x

Deshpande, K., “N-colour separation methods for accurate reproduction of spot colours”, PhD Thesis, University of the Arts London, London, United Kingdom, 2015.

Dharavath, Nagaraju & Hahn, Kim. (2009). Green Printing: Colorimetric and Densitometric Analysis of Solvent-based and Vegetable Oil-based Inks of Multicolor Offset Printing. The Journal of Technology Studies. 35. 10.21061/jots.v35i2.a.4.

Dobrusskin,C., On the Identification of Contradictions Using Cause Effect Chain Analysis, Procedia CIRP, Volume 39, 2016, Pages 221-224, ISSN 2212-8271,https://doi.org/10.1016/j.procir.2016.01.192.

Harris, P. and Moore, P. (2004). METHODS OF FLEXOGRAPHIC PRINTING WITH INKS EXHIBITING EXPANDED COLOR-SPACE CHARACTERISTICS. US 6834589.

Hirschler, R., Csillag, P., Manyé, P., Neder, M., 2018. How much colour science is not too much? Color Research and Application 43, 977–992. https://doi.org/10.1002/col.22275

Idealliance®, 2021. Disponible en: https://idealliance.org/specifications/gracol/

Izdebska, J. (2016). Printing on Polymers. Printing on Polymers, 1–20. doi:10.1016/b978-0-323-37468-2.00001-4

Gamm, B. M., & Gerard, P. (2020). Testing the differences between two color measurement probability distributions using Hotelling's T 2 test and the permutation test. Color Research & Application, 45(2), 196–207. doi:10.1002/col.22468

Gutierrez-Pulido, H., Gutierrez-Gonzales, P., Garibay-Lopez, C., y Diaz-Caldera, L., Análisis multivariado y QFD como herramientas para escuchar la voz del cliente y mejorar la calidad del servicio. Ingeniare. Rev. chil. ing. [online]. 2014, vol.22, n.1, pp.62-73. ISSN 0718-3305. http://dx.doi.org/10.4067/S0718-33052014000100007.

Jangra, V., Saini, A., & Kundu, A. (2014). Relationship of Solid Ink Density and Dot Gain in Digital Printing. International Journal of Engineering and Technical Research (IJETR), 2321-0869.

Korifi, R., Le Dréau, Y., Antinelli, J.-F., Valls, R., & Dupuy, N. (2013). CIEL∗a∗b∗ color space predictive models for colorimetry devices – Analysisof perfume quality. Talanta, 104, 58–66. doi:10.1016/j.talanta.2012.11.026

Labrecque, L. I. (2020). Color research in marketing: Theoretical and technical considerations for conducting rigorous and impactful color research. Psychology & Marketing, 37(7), 855–863. doi:10.1002/mar.21359

León, K., Mery, D., Pedreschi, F., & León, J. (2006). Color measurement in L∗a∗b∗ units from RGB digital images. Food Research International, 39(10), 1084–1091. doi:10.1016/j.foodres.2006.03.006

Mahajan, M. P., & Bandyopadhyay, S. (2019). Characterization and optimization of color attributes chroma (C *) and lightness (L *) in offset lithography halftone print on packaging boards. Color Research & Application. doi:10.1002/col.22456

Meyn, J.-P. (2008). Colour mixing based on daylight. European Journal of Physics, 29(5), 1017–1031. doi:10.1088/0143-0807/29/5/014

Neyestani, B.,, Seven Basic Tools of Quality Control: The Appropriate Techniques for Solving Quality Problems in the Organizations (2017). Available at SSRN: https://ssrn.com/abstract=2955721 or http://dx.doi.org/10.2139/ssrn.2955721

Olmedo- García L. R. (2009) PROPUESTA PARA LA PRUEBA DE UNA HIPÓTESIS NULA CENTRAL COMPUESTA Y UNA HIPÓTESIS ALTERNATIVA BILATERAL EN LA DISTRIBUCIÓN NORMAL [Tesis de Maestría, Universidad Autónoma Metropolitana].

Patel, S., "Determining the effect of printing ink sequence for process colors on color gamut and print quality in flexography", M.S. Thesis. Rochester Institute of Technology, Henrietta, NY, USA, 2009.

Peña, Daniel. (2002). Análisis de Datos Multivariantes.

Prust, Z. (2010). Graphic Communications. 5th ed. p.383.

Rosato, D. V. (Ed.). (2000). Concise encyclopedia of plastics. New York: Springer Science & Business Media.

Sharma, A. (2018) ‘Spot Colors & Expanded Gamut Printing’, Understanding Color Management, pp. 233–258. doi: 10.1002/9781119223702.ch9.

Sharma A, Seymour J. Evaluation of expanded gamut software solutions for spot color reproduction. Color Res Appl. 2019; 1–10. https://doi.org/10.1002/col.22471

Sheth, "Extended Color Gamut for Flexographic Printing" (2013). Master's Theses. 138. https://scholarworks.wmich.edu/masters_theses/138

Silayoi, P., & Speece, M. (2007). The importance of packaging attributes: a conjoint analysis approach. European Journal of Marketing, 41(11/12), 1495–1517. doi:10.1108/03090560710821279

Tutak, Dogan & Beytut, Huseyin & Ozcan, Arif. (2018). Investigation of the effects of different ink density values on color gamut in offset printing. Journal of Graphic Engineering and Design. 9. 23-28. 10.24867/JGED-2018-1-023.

Verikas, A., & Bacauskiene, M. (2008). Estimating ink density from colour camera RGB values by the local kernel ridge regression. Engineering Applications of Artificial Intelligence, 21(1), 35–42.

X-Rite, 2020. Disponible en: https://www.xrite.com/es/blog/densitometer-density-measurement

Zhao, X. (2007) ‘Generic device color gamut description’, Color Research and Application, 32(5), pp. 394–408. doi: 10.1002/col.20343.

7. ANEXOS

Figura 7.1 Proceso de registro en impresión

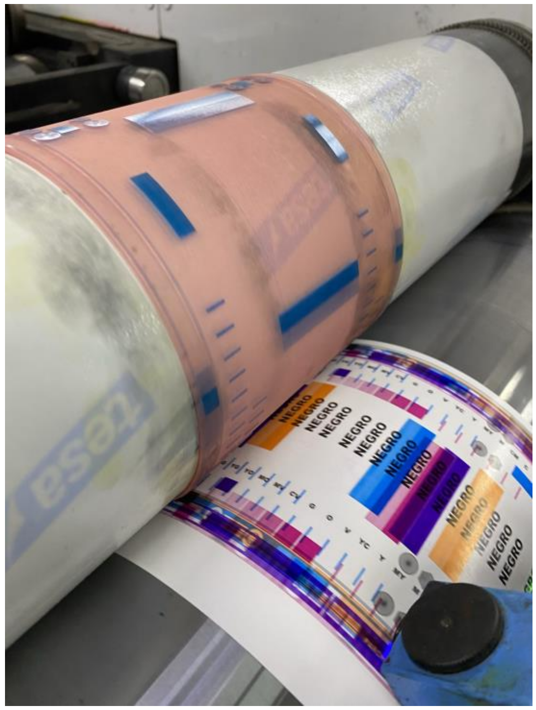

Figura 7.2 Impresión siendo registrada

Figura 7.3 Lado izquierdo impresión de máquina 7 y lado derecho impresión de máquina 5

```matlab
n= 60;
n1=60;
n2 =60;
p= 4;
magentadeseado = [48.12 74.15 -5.05];
LdeseadoM = magentadeseado(:,1);
adeseadoM = magentadeseado(:,2);
bdeseadoM = magentadeseado(:,3);

%%%%%%%%%% ESTIMACIÓN DE PARÁMETROS P5 MAGENTA NAZDAR %%%%%%
MagentaP5Nazdar = 'P5MagentaNazdar.xlsx';
datosMagentaP5N = xlsread(MagentaP5Nazdar);
mediaMagentaP5Nazdar = mean(datosMagentaP5N);
varianzaMagentaP5Nazdar = cov(datosMagentaP5N);
correlacionMagentaP5N = corrcoef(datosMagentaP5N);
precisionMagentaP5N = inv(varianzaMagentaP5Nazdar);

filaMagentaP5N = precisionMagentaP5N(4,:);

valordensidadprecisionP5MN = filaMagentaP5N(:,4);
residualprevioP5MN = filaMagentaP5N/valordensidadprecisionP5MN;

LresidualprevioP5MN = residualprevioP5MN(:,1);
LmediaP5MN = mediaMagentaP5Nazdar(:,1);

aresidualprevioP5MN = residualprevioP5MN(:,2);
amediaP5MN = mediaMagentaP5Nazdar(:,2);

bresidualprevioP5MN = residualprevioP5MN(:,3);
bmediaP5MN = mediaMagentaP5Nazdar(:,3);

dresidualprevioP5MN = residualprevioP5MN(:,4);
dmediaP5MN = mediaMagentaP5Nazdar(:,4);

densidaddeseadaP5MN = dmediaP5MN-(LresidualprevioP5MN*(LdeseadoM-LmediaP5MN))-(aresidualprevioP5MN*(adeseadoM-amediaP5MN))-(bresidualprevioP5MN*(bdeseadoM-bmediaP5MN));
valoresdeseadosP5MN = [LdeseadoM,adeseadoM,bdeseadoM,densidaddeseadaP5MN];
%%%prueba de hipotesis comparando medias de valores deseados %%%
mediaHipotesisP5MN = valoresdeseadosP5MN;
diferenciademediasM5N = mediaMagentaP5Nazdar - mediaHipotesisP5MN;

TcuadP5MN = n1*(mediaMagentaP5Nazdar - mediaHipotesisP5MN) * precisionMagentaP5N*(mediaMagentaP5Nazdar - mediaHipotesisP5MN)';
FP5MN = ((n1-p)/(p*(n1-1)))*TcuadP5MN;
valorpruebaP5MN = 1-fcdf(FP5MN,4,56);


%%%%%%%%%% ESTIMACIÓN DE PARÁMETROS P7 MAGENTA NAZDAR %%%%%%
MagentaP7Nazdar = 'P7MagentaNazdar.xlsx';
datosMagentaP7N = xlsread(MagentaP7Nazdar);
mediaMagentaP7Nazdar = mean(datosMagentaP7N);
varianzaMagentaP7Nazdar = cov(datosMagentaP7N);
correlacionMagentaP7N = corrcoef(datosMagentaP7N);
precisionMagentaP7N = inv(varianzaMagentaP7Nazdar);

filaMagentaP7N = precisionMagentaP7N(4,:);

valordensidadprecisionP7MN = filaMagentaP7N(:,4);
residualprevioP7MN = filaMagentaP7N/valordensidadprecisionP7MN;

LresidualprevioP7MN = residualprevioP7MN(:,1);
LmediaP7MN = mediaMagentaP7Nazdar(:,1);

aresidualprevioP7MN = residualprevioP7MN(:,2);
amediaP7MN = mediaMagentaP7Nazdar(:,2);

bresidualprevioP7MN = residualprevioP7MN(:,3);
bmediaP7MN = mediaMagentaP7Nazdar(:,3);

dresidualprevioP7MN = residualprevioP7MN(:,4);
dmediaP7MN = mediaMagentaP7Nazdar(:,4);

densidaddeseadaP7MN = dmediaP7MN-(LresidualprevioP7MN*(LdeseadoM-LmediaP7MN))-(aresidualprevioP7MN*(adeseadoM-amediaP7MN))-(bresidualprevioP7MN*(bdeseadoM-bmediaP7MN));
valoresdeseadosP7MN = [LdeseadoM,adeseadoM,bdeseadoM,densidaddeseadaP7MN];
%%%prueba de hipotesis comparando medias de valores deseados %%%
mediaHipotesisP7MN = valoresdeseadosP7MN;
diferenciademediasM7N = mediaMagentaP7Nazdar - mediaHipotesisP7MN;

TcuadP7MN = n1*(mediaMagentaP7Nazdar - mediaHipotesisP7MN) * precisionMagentaP7N*(mediaMagentaP7Nazdar - mediaHipotesisP7MN)';
FP7MN = ((n1-p)/(p*(n1-1)))*TcuadP7MN;
valorpruebaP7MN = 1-fcdf(FP7MN,4,56);

%%%%%% PRUEBA DE HIPOTESIS uP5 = uP7  %%%
sMN = ((59*varianzaMagentaP5Nazdar)+(59*varianzaMagentaP7Nazdar))/(n1+n2-2);
swMN = varianzaMagentaP7Nazdar;
lambdaMN = n*log(det(sMN)/det(swMN));
valorMN = 1- chi2cdf(lambdaMN,4);
```

Figura 7.4 Código utilizado en MATLAB