

MÉTODOS ESTADÍSTICOS PARA EVALUAR LA CALIDAD DEL CAFÉ

Claudia Alejandra SALAMANCA ROMERO

Dipòsit legal: Gi. 2065-2015
<http://hdl.handle.net/10803/327037>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

**Tesis
Doctoral**



MÉTODOS ESTADÍSTICOS PARA EVALUAR LA CALIDAD DEL CAFÉ

**CLAUDIA ALEJANDRA
SALAMANCA ROMERO**

2015





Universitat de Girona

TESIS DOCTORAL

**MÉTODOS ESTADÍSTICOS PARA EVALUAR LA
CALIDAD DEL CAFÉ**

Claudia Alejandra Salamanca Romero

2015



Universitat de Girona

TESIS DOCTORAL

**MÉTODOS ESTADÍSTICOS PARA EVALUAR LA
CALIDAD DEL CAFÉ**

Claudia Alejandra Salamanca Romero

2015

**Programa De Doctorado En Ciencias Experimentales y
Sostenibilidad**

Directores:

Dra. Isabel Villaescusa Gil

Dr. Marc Saez Zafra

Memoria presentada para optar al título de doctora por la Universidad de Girona



Universitat de Girona

La Dra. Isabel VILLAESCUSA GIL y el Dr. Marc SAEZ ZAFRA, de la universidad de Girona,

DECLARAMOS:

Que hemos supervisado el trabajo realizado por Claudia Alejandra SALAMANCA ROMERO titulado “Métodos Estadísticos Para Evaluar La Calidad Del Café” y que se presenta en formato de monografía tradicional para solicitar el título de Doctor por la Universidad de Girona.

Firma,

Dra. Isabel VILLAESCUSA GIL

Dr. Marc SAEZ ZAFRA

Girona, febrero de 2015

A mi familia que sacrificó su tiempo para que yo pudiera cumplir con el mío. Hoy en día puedo decir que esta tesis lleva mucho de ustedes.

AGRADECIMIENTOS

Son muchas las personas a las que debo agradecer por respaldar de una u otra manera este proyecto.

A la Profesora Dra. Isabel VILLAESCUSA, Al Profesor Dr. Marc SAEZ y al director comercial Carles GONZALEZ de la empresa Rancilio España S.A. por la orientación, el seguimiento, la motivación y el apoyo continuo recibido durante estos años, pero sobre todo, por confiar en mí y por hacer posible mi sueño.

También quiero dar las gracias a la Dra. Núria FIOL, Dr. Florencio DE LA TORRE, a mis compañeros: María PEIROLO, David PUJOL, Chang LIU y al personal técnico del laboratorio: Nuria CABEZAS y Gemma GIBERT por su colaboración en algún momento del desarrollo de esta investigación.

Un agradecimiento muy especial a mi hija Andréa y a mi esposo Frédéric QUINTANA por su comprensión y paciencia.

A todos ellos, muchas gracias.

LISTA DE FIGURAS

Figura 1. Procesos de beneficio, tostado y preparación de la bebida del café.	2
Figura 2. Estructura de la tesis.	48
Figura 3. Mapa del Valle del Cauca con sus municipios cafeteros incluidos en el estudio.	50
Figura 4. Maquina Xcelsius clase 9 de la empresa Rancilio.	57
Figura 5. Variables antes y después de los procesos de imputación.	71
Figura 6. Relación entre las variables iniciales.	72
Figura 7. Eficiencia Relativa de la estimación a través de imputación múltiple según el número de imputaciones (m) y la fracción de información faltante (λ).	74
Figura 8. Variables antes y después del proceso de imputación múltiple MI.	76
Figura 9. Mapa del Departamento del Valle del Cauca que presenta una media, buena y excelente calidad.	81
Figura 10. Variedades de café según el Índice de calidad.	82
Figura 11. Certificación de café según el Índice de calidad.	83
Figura 12. Puntuación sensorial del café según el Índice de calidad.	84
Figura 13. Defectos presentes en las pruebas de taza contrastados con el índice de calidad.	84
Figura 14. Distribución del tamaño de partículas según El tipo de Café y la temperatura.	88
Figura 15. Propiedades Físicas del Café tipo A (Robusta Natural).	93
Figura 16. Propiedades Físicas del Café Tipo B (Arábica Natural).	94
Figura 17. Propiedades Físicas del Café Tipo C (Arábica Lavado).	95

Figura 18. Índice de crema por café y por temperatura.	96
Figura 19. Densidad por café y por temperatura.	97
Figura 20. Turbidez, Sólidos Totales, Sólidos filtrados y Viscosidad por café y por temperatura.	99
Figura 21. Propiedades químicas del Café Tipo A (Robusta natural).	100
Figura 22. Propiedades químicas del Café Tipo B (Arábica natural).	101
Figura 23. Propiedades químicas del Café Tipo C (Arábica Lavado).	101
Figura 24. pH por tipo de café y por temperatura.	103
Figura 25. Polifenoles, Acido Clorogénico y Cafeína por tipo de café y por temperatura.	105
Figura 26. Lípidos por tipo de café y por temperatura.	107
Figura 27. Correlaciones entre las propiedades fisicoquímicas y los descriptores organolépticos en el café Robusta Natural.	112
Figura 28. Correlaciones entre los descriptores organolépticos y las propiedades fisicoquímicas en el café Arábica Natural.	114
Figura 29. Correlaciones entre los descriptores organolépticos y las propiedades fisicoquímicas en café Arábica Lavado.	115
Figura 30. Primer plano factorial para el perfil de temperatura ascendente	120
Figura 31. Primer plano factorial para el perfil de temperatura ascendente	122
Figura 32. Primer plano factorial para el perfil de temperatura descendente	124

LISTA DE TABLAS

Tabla 1. Ventajas y desventajas de los métodos de imputación CD, HD y MI.	29
Tabla 2. Variables del análisis sensorial de las pruebas de taza que son resumidas en la puntuación final.	51
Tabla 3. Variables de análisis físico de calidad de café.	53
Tabla 4. Variables Espaciales y Agronómicas del estudio.	54
Tabla 5. Calidad de Agua.	56
Tabla 6. Clasificación del tipo de café	57
Tabla 7. Descriptores organolépticos	65
Tabla 8. Variables de los datos originales (sin imputación).	67
Tabla 9. Variables imputadas bajo las diferentes técnicas.	68
Tabla 10. Eficiencia de la estimación a través de imputación múltiple según el número de imputaciones (m) y la fracción de información faltante (λ).	74
Tabla 11. Factor de cargas (λ^2) y varianza del Error (ψ_{jj}).	78
Tabla 12. Frecuencias para el Índice de calidad – IC (ϕ_i).	79
Tabla 13. Municipios del Valle del Cauca que forman parte del estudio.	81
Tabla 14. Tamaño de partícula y clasificación de la molienda.	87
Tabla 15. Estadísticas descriptivas de las variables físicas ($n = 18$).	90
Tabla 16. Estadísticas descriptivas de las variables químicas ($n = 18$).	91
Tabla 17. Media, desviación estándar y Coeficiente de variación de los parámetros físicos químicos para los perfiles o curvas de temperatura.	111
Tabla 18. Índice de Bartlett e índice de Kaiser-Meyer-Olkin- KMO.	117
Tabla 19. Proporción de variabilidad explicada por componente.	118
Tabla 20. Comunalidades por perfiles de temperatura.	118

Tabla 21. Componentes principales en el perfil de temperatura ascendente.	119
Tabla 22. Componentes principales en el perfil de temperatura Plana.	121
Tabla 23. Componentes principales en el perfil de temperatura Descendente.	123

SUMMARY

This thesis was devoted to the analysis and evaluation of statistical methods to assess the quality of the coffee. In this sense, a contribution to try to quantify and qualify the coffee quality through statistical techniques addressed from Bayesian statistics, multivariate and experimental design were given. The methods described in each of the studies carried out, show the relationship and the differences of the different techniques used.

In the first study, the coffee index of quality of the coffee from the department of Valle del Cauca in Colombia was developed taking into account the organoleptic and agronomic characteristics using Bayesian factor analysis. A database with some missing values (missing values). containing different organoleptic tests and agronomic aspects of the crop was used in the analysis. For the treatment of the data it was necessary to use data imputation techniques (listwise or case deletion (CD) Hot deck (HD) and multiple imputation (MI) of which database product chosen multiple imputation (MI), since this technique showed more goodness in relation to others. Finally, it was found for the first time, coffee index of quality using a multiple factor analysis of Bayesian character due to the nature of the data as possible and complex parameterization. This index reflected spatially coffee crop areas that are identified as areas that produce good quality coffee. Besides other analyzes of interest in relation to the variety, profit, certification and others were performed.

In the second study, the quality of the espresso was determined by investigating the effect of extraction temperature on the physicochemical properties of espresso, using three types of Natural Robusta coffees (Type A), Natural Arabica (Type B) and washed Arabica (Type C) and three different temperature profiles: upward extraction temperature (88 - 93 °C), flat (90-90 °C) and downstream (93-88 °C) with the latest

innovation machine XCELSIUS RANCILIO class 9 trademark. The experimental design statistical tool known as the ANOVA analysis of variance of two factors (type of coffee and temperature profile) was used. Analyses showed that the use of temperature profiles allows modifying, modulate, enhance or reduce the organoleptic descriptors of an espresso.

Finally, **in the third study**, the discriminatory elements to characterize espresso using the three types of coffees and profiles or extraction temperature curves that were used in the previous study were identified. The physicochemical properties were determined and cup tests were performed through a tasting panel that yielded organoleptic descriptors for each CE. The statistical processing using the multivariate technique called principal component analysis (PCA) was performed to analyze and visualize possible correlations, and to allow the identification of trends in the quality and their relationship to physicochemical properties.

RESUM

En aquesta tesi es va realitzar l'anàlisi i l'avaluació dels mètodes estadístics per a valorar la qualitat del cafè. En aquest sentit, es va tractar de quantificar i qualificar la qualitat del cafè mitjançant tècniques estadístiques abordades des de l'estadística Bayesiana multivariada i del disseny experimental. Els mètodes que es descriuen en cadascun dels estudis mostren la relació i les diferències de les distintes tècniques emprades.

En el **primer estudi** es va elaborar l'índex de qualitat del cafè del 'Departamento del Valle del Cauca', a Colòmbia, a partir de les característiques organolèptiques i agronòmiques, utilitzant anàlisi factorial Bayesià. Es va partir d'una base de dades que contenia diferents proves de tassa i els aspectes agronòmics del cultiu amb alguns valors mancants (missing values). Per al seu maneig i depuració va ser necessari utilitzar tècniques d'imputació de dades (listwise o case deletion (CD), Hot Dect (UHD) i imputació múltiple (MI) de les quals es va escollir la base de dades producte de la imputació múltiple, ja que aquesta va ser la que va mostrar més bonesa en relació a les altres. Finalment, es va trobar, per primer cop, l'índex de qualitat del cafè, utilitzant un Anàlisi Factorial Múltiple de caràcter Bayesià, degut a la natura de les dades i a una possible i complexa parametrització. Amb aquest índex, es van caracteritzar espacialment les àrees cafeteres del Departament, productores d'un cafè de qualitat. A més, es van realitzar altres anàlisis d'interès amb relació a la varietat, benefici, certificació, etc., entre d'altres.

En el **segon estudi** es va determinar la qualitat del cafè expresso mitjançant la investigació de l'efecte de la temperatura d'extracció en les propietats físico-químiques del cafè expresso, utilitzant tres tipus de cafè: Robusta Natural (Tipus A), Aràbica Natural (Tipus B) i Aràbica Rentat (Tipus C) i tres diferents perfils o corbes

de temperatura d'extracció: ascendent (88–93 °C), plana (90-90 °C) i descendent (93–88 °C) amb la màquina de darrera innovació XCELSIUS classe 9 de la marca RANCILIO. Com a eina estadística es va utilitzar el disseny experimental ANOVA conegut com l'anàlisi de la variància de dos factors (tipus de cafè i perfil de temperatura). Els anàlisis realitzats van demostrar que la utilització dels perfils o corbes de temperatura van permetre modificar les propietats físico-químiques i modular, exaltar o rebaixar els descriptors organolèptics d'un expresso.

Finalment, el **tercer estudi**, es va identificar els elements discriminatoris per a caracteritzar el cafè expresso, utilitzant els tres tipus de cafès i perfils o corbes de temperatura d'extracció que es van utilitzar en el segon estudi. Es van determinar les propietats físico-químiques i es van realitzar proves de tassa mitjançant un panell de tast que va permetre obtenir els descriptors organolèptics de cada CE. El processament estadístic utilitzat va ser l'Anàlisi de Components Principals (ACP), que va permetre analitzar i visualitzar correlacions, a més de permetre la identificació de les tendències en la qualitat i la seva relació amb les propietats físico-químiques.

RESUMEN

La presente tesis realizó el análisis y evaluación de los métodos estadísticos para valorar la calidad del café. En este sentido, se dio un aporte para tratar de cuantificar y cualificar la calidad del café mediante las técnicas estadísticas abordados desde la estadística bayesiana, multivariada y del diseño experimental. Los métodos que se describen en cada uno de los estudios, muestran la relación y las diferencias de las diferentes técnicas empleadas.

En el **primer estudio**, se elaboró el índice de calidad del café del departamento del valle del cauca en Colombia a partir de las características organolépticas y agronómicas empleando análisis factorial bayesiano. Se partió de una base de datos que contenía diferentes pruebas de taza y los aspectos agronómicos del cultivo con algunos valores faltantes (missing values). Para su manejo y depuración, fue necesario utilizar técnicas de imputación de datos (listwise o case deletion (CD), Hot deck (HD) e imputación múltiple (MI) de las cuales se escogió la base de datos producto de la imputación múltiple (MI), ya que esta técnica que mostró más bondades con relación a las demás. Finalmente, se encontró por primera vez, el índice de calidad del café, usando un Análisis Factorial múltiple de carácter bayesiano debido a la naturaleza de los datos y a una posible y compleja parametrización. Con este índice, se reflejó espacialmente las aéreas cafeteras del departamento que se caracterizaron por producir un café de calidad. Además se realizaron otros análisis de interés con relación a la variedad, beneficio, certificación entre otros.

En el **segundo estudio**, se determinó la calidad del café espresso mediante la investigación del efecto de la temperatura de extracción en las propiedades fisicoquímicas del café espresso, utilizando tres tipos de cafés Robusta Natural (Tipo

A), Arábica Natural (Tipo B) y Arábica Lavado (Tipo C) y tres diferentes perfiles o curvas de temperatura de extracción ascendente (88 – 93 °C), plana (90-90 °C) y descendente (93–88 °C) con la máquina de última innovación XCELSIUS clase 9 de la marca RANCILIO. Como herramienta estadística se utilizó el diseño experimental ANOVA conocido como el análisis de varianza de dos factores (Tipo de café y perfil de temperatura). Los análisis realizados demostraron que la utilización de los perfiles o curvas de temperatura permitieron modificar las propiedades fisicoquímicas y modular, exaltar o rebajar los descriptores organolépticos de un espresso.

Finalmente en el **tercer estudio**, se identificó los elementos discriminatorios para caracterizar el café espresso empleando los tres tipos de cafés y perfiles o curvas de temperatura de extracción que se usaron en el estudio anterior. Se determinaron las propiedades fisicoquímicas y se realizaron pruebas de taza a través de un panel de cata que permitió obtener los descriptores organolépticos de cada CE. El procesamiento estadístico realizado se hizo mediante la técnica multivariada denominada análisis de componentes principales (ACP) que permitió analizar y visualizar correlaciones, además de permitir la identificación de las tendencias en la calidad y su relación con las propiedades fisicoquímicas.

TABLA DE CONTENIDO

AGRADECIMIENTOS	i
LISTA DE FIGURAS	iii
LISTA DE TABLAS	v
SUMMARY	vii
RESUM	ix
RESUMEN	xi
1. INTRODUCCIÓN GENERAL	1
1.1. <i>EL CAFÉ</i>	1
1.2. <i>CALIDAD DEL CAFÉ</i>	6
1.3. <i>LA BEBIDA DEL CAFÉ</i>	8
1.4 <i>CALIDAD DEL CAFÉ ESPRESSO</i>	9
1.4.1 <i>ASPECTOS QUE DETERMINAN LA CALIDAD DE UN ESPRESSO</i>	10
<i>Calidad del agua</i>	10
<i>Tamaño de partícula - granulometría</i>	11
<i>Crema</i>	12
<i>Temperatura</i>	12
1.5. <i>MÉTODOS ESTADÍSTICOS</i>	12
1.5.1 <i>DATOS FALTANTES (MISSING VALUES)</i>	13
1.5.1.1. <i>MECANISMO DE PÉRDIDA DE LOS DATOS FALTANTES</i>	13

1.5.1.2. MÉTODOS PARA EL MANEJO DE DATOS FALTANTES	16
MÉTODOS CONVENCIONALES	16
ListWise (Case Deletion) o eliminación por listas	16
PairWise o Eliminación por pareja	16
MÉTODO DE IMPUTACIÓN SIMPLE	17
Hot-Deck o fichero caliente	17
MÉTODO DE IMPUTACIÓN MÚLTIPLE	18
Imputación Múltiple por Ecuaciones Encadenadas – MICE	23
1.5.1.3 DISCUSIÓN DE LOS MÉTODOS DE IMPUTACIÓN	26
1.6. ANÁLISIS FACTORIAL	30
Análisis de la matriz de correlación	31
Extracción factorial	32
La fase de rotación	33
Cálculo de puntuaciones factoriales	34
1.6.1. ANÁLISIS FACTORIAL BAYESIANO	35
Modelo de análisis factorial para datos mixtos	37
Especificación a priori del modelo completo	39
Modelo de ajuste e inferencia	40
Diagnóstico convergencia de MCMC	41
1.7 ANÁLISIS DE VARIANZA	42
Pruebas Post- Anova	43
1.8. PLANTEAMIENTO GENERAL	44
2. OBJETIVOS	46
3. METODOLOGÍA	48
<i>Introducción</i>	48
3.1. ÍNDICE DE CALIDAD DEL CAFÉ DEL VALLE DEL CAUCA EN COLOMBIA A PARTIR DE LAS CARACTERÍSTICAS ORGANOLÉPTICAS Y AGRONÓMICAS	49
<i>Ubicación geográfica del estudio</i>	49

<i>Obtención de la Información</i>	50
<i>Características de las variables</i>	51
<i>Descripción del método estadístico</i>	54
3.2. CALIDAD DEL CAFÉ ESPRESSO A PARTIR DE LAS PROPIEDADES FÍSICOQUÍMICAS EMPLEANDO PERFILES DE TEMPERATURA DE EXTRACCIÓN	55
<i>Establecimiento de las condiciones de laboratorio</i>	56
<i>Procedimiento de extracción y calibración de la maquina</i>	57
<i>Preparación de la muestra</i>	57
<i>Propiedades físicas y químicas del café espresso – (CE)</i>	58
<i>Granulometría o tamaño de partícula</i>	61
<i>Descripción del diseño y análisis del método estadístico</i>	61
3.3. ELEMENTOS DISCRIMINATORIOS PARA CARACTERIZAR EL CAFÉ ESPRESSO	62
<i>Obtención de los cafés espressos</i>	63
<i>Variables del estudio</i>	63
<i>Descripción del método estadístico</i>	65
4. RESULTADOS Y DISCUSIÓN	67
4.1 ÍNDICE DE CALIDAD DEL CAFÉ DEL VALLE DEL CAUCA EN COLOMBIA A PARTIR DE LAS CARACTERÍSTICAS ORGANOLÉPTICAS Y AGRONÓMICAS.	67
4.1.1 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES	67
4.1.2 ANÁLISIS DE LOS RESULTADOS PARA VALORES FALTANTES	73
4.1.3 ANÁLISIS FACTORIAL BAYESIANO	76
4.1.4 USO DE LA IMPUTACIÓN MÚLTIPLE Y ANÁLISIS FACTORIAL BAYESIANO	85
4.2 CALIDAD DEL CAFÉ ESPRESSO A PARTIR DE LAS CUALIDADES	

<i>FISICOQUÍMICAS</i>	87
<i>4.2.1 GRANULOMETRÍA</i>	87
<i>4.2.2 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES</i>	89
<i>4.2.3 PROPIEDADES FÍSICAS POR CAFÉ OBTENIDOS A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN</i>	93
<i>4.2.4 COMPARACIÓN DE ALGUNAS PROPIEDADES FÍSICAS DEL CAFÉ A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN</i>	95
<i>4.2.5 PROPIEDADES QUÍMICAS POR CAFÉ A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN</i>	99
<i>4.2.6 COMPARACIÓN DE ALGUNAS PROPIEDADES QUÍMICAS DEL CAFÉ A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN</i>	102
<i>4.2.7. RELACIÓN DE LAS PROPIEDADES FISICOQUÍMICAS CON LAS PROPIEDADES ORGANOLÉPTICAS</i>	108
<i>4.3. ELEMENTOS DISCRIMINATORIOS PARA CARACTERIZAR EL CAFÉ ESPRESSO A TRAVÉS DE DESCRIPTORES ORGANOLÉPTICOS Y PROPIEDADES FISICOQUÍMICAS</i>	109
<i>4.3.1 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES</i>	109
<i>4.3.2 RELACION ENTRE LOS DESCRIPTORES ORGANOLEPTICOS Y LAS PROPIEDADES FISICOQUIMICAS</i>	112
<i>4.3.3 ANÁLISIS DE COMPONENTES PRINCIPALES – ACP POR PERFIL DE TEMPERATURA</i>	116
<i>4.3.4. RELACIÓN DE LAS PROPIEDADES FISICOQUÍMICAS Y ORGANOLÉPTICAS CON LA TEMPERATURA DE EXTRACCIÓN</i>	124
5. CONCLUSIONES	126
BIBLIOGRAFIA	129

1. INTRODUCCIÓN GENERAL

1.1. EL CAFÉ

El cafeto es un arbusto perenne que pertenece a la familia de las rubiáceas y del género *Coffea*. Son arbustos que tienen una altura entre los 10-12 metros, pero para facilitar su manejo, se poda entre los 2 y 4 metros, además son plantas que requieren de climas y suelos tropicales. La primera cosecha del café ocurre a los dos años pero solo después de los cuatro o cinco años comienzan a recogerse las cosechas óptimas. El fruto del café es una cereza que cuando nace es de un color verde y luego cambia a amarillo o hasta tomar un color rojo, lo que significa que ha alcanzado su plena madurez, en el interior de cada cereza, hay dos semillas separadas por un surco y rodeadas de una pulpa amarilla. Estos granos están protegidos por una película plateada llamada mucilago y recubiertos por una piel de color amarillo (Arcila et al. 2007).

En el comercio, los cafés se especifican según la especie (Arábica o Robusta) y según el tipo de beneficio (húmedo o seco) que es el proceso conformado por un conjunto de operaciones para transformar el café cereza en café pergamino seco (CPS). El beneficio del grano por la vía seca, es el secado directo de los frutos de café en el árbol o en secaderos después del despulpado. En el beneficio por la vía húmeda, después del despulpado, el mucílago del fruto es separado mediante fermentación, lavado y clasificado del café, luego el grano es secado para reducir la humedad hasta llegar entre el 10 y el 12% como norma de comercialización del CPS. Según el beneficio, los granos almendra presentan diferencias: los cafés Arábica por vía húmeda son verdes, mientras que por vía seca son amarillos o marrones (Puertas, 2011a).

En el diagrama de la Figura 1, se muestran los diferentes procesos que sufre el café

desde su recolección hasta la preparación de la bebida.

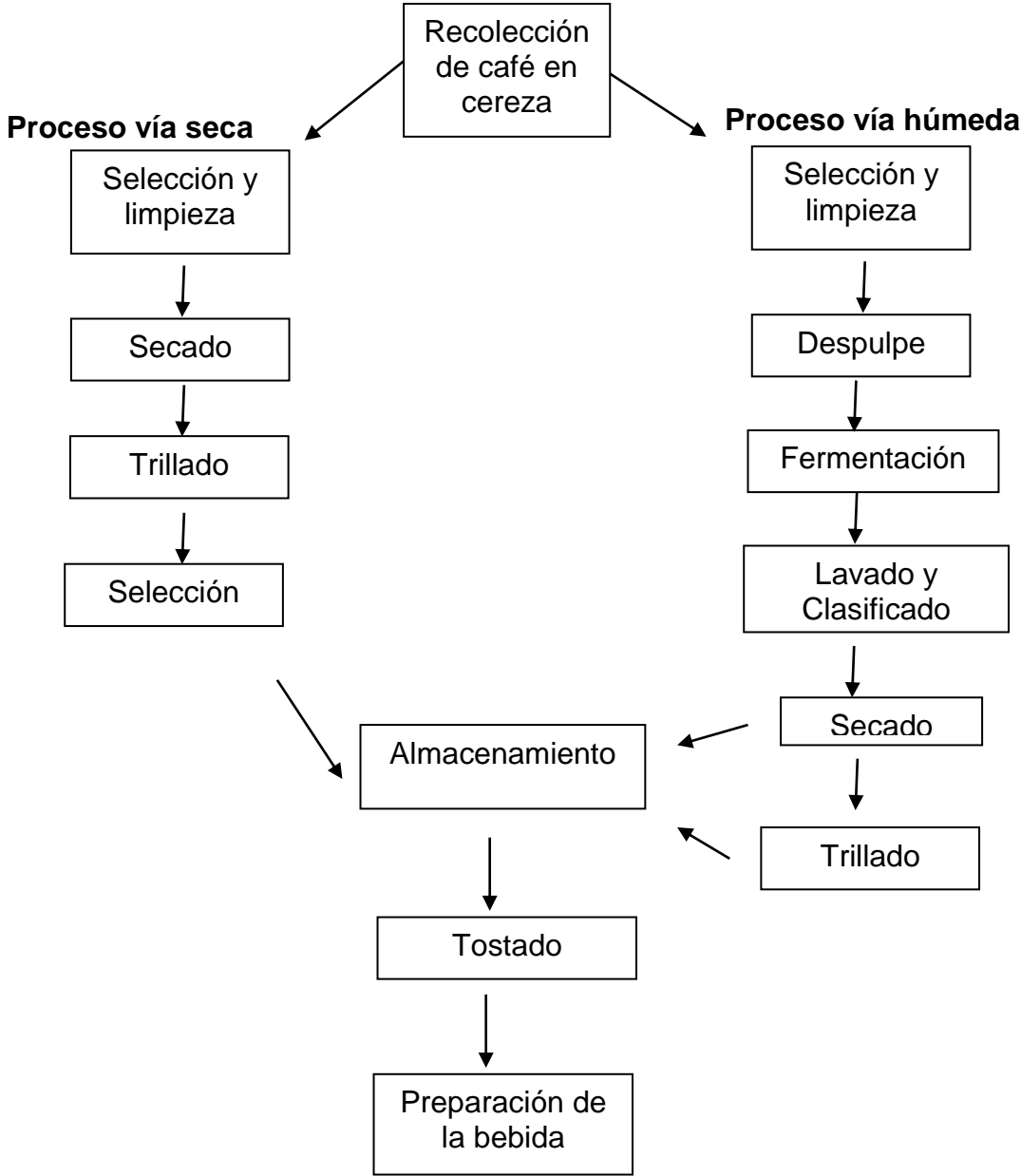


Figura 1. Procesos de beneficio, tostado y preparación de la bebida del café.

Cabe destacar que en esta tesis se ha partido del café después del proceso de tostado para realizar todos los experimentos.

El cultivo de café por su importancia económica y social se ha catalogado como uno de los principales productos agrícolas de Colombia y del mundo. Hoy en día se reconocen alrededor de 103 especies de café pertenecientes a la familia de las Rubiáceas, pero sólo dos son responsables del 99% del comercio mundial, la ***Coffea canephora*** con una participación del 35% (conocida también como *Coffea Robusta*) distribuida geográficamente en África, en países como en el Congo, Sudán, Uganda, el Noroeste de Tanzania y Angola. La ***Coffea Arábica*** con el porcentaje restante, es la más apreciada debido a su bajo contenido de cafeína. Se produce principalmente en países del centro y sur de América como: Costa Rica, Ecuador, Brasil, Colombia, Perú, entre muchos otros.

El cuanto a la comercialización del café, la OIC¹ informó que el consumo mundial en el 2013 subió a 146 millones de sacos frente a los 142 millones del año anterior. Para el año 2012, el mayor país consumidor en las Américas, fue EE.UU con 22.2 millones de sacos lo que representa el 15.7% del consumo mundial, frente a la UE donde el consumo total fue de 41 millones de sacos. En Alemania el consumo fue de 9.5 millones y en Francia de 5.4 millones de sacos, mostrando descensos respectivos del 6.7% y 2.9%. Por otro lado, España tuvo un notable aumento del 9.1% con 3.5 millones de sacos. El consumo de café en Italia (5.7 millones de sacos) y el Reino Unido (2.8 millones de sacos) se mantuvo relativamente igual con respecto al 2011. No obstante, el consumo de café es muy variable y depende en gran medida de la forma como se consume y de su calidad.

En una economía como la actual, las necesidades de obtener productos de calidad que se adecuen a normas, exigencias y estándares de calidad son la base de la comercialización de cualquier producto. En el caso del café, muchos autores tratan de explicar la calidad desde el punto de vista de su disciplina particular. **En el sistema de producción**, Benítez et al. (2003) y Tobasura (2006) mencionan que la apreciación de calidad, solo se ha buscado mediante la identificación y fortalecimiento del cultivo a través de los cafés especiales, bajo criterios de sostenibilidad. Para Wang (2012), la calidad del grano del café es el resultado de los procesos de cosecha y poscosecha por que influye significativamente en su

¹ Organización Internacional del Café OIC 2013

composición química. **En los procesos de comercialización del café**, Fermín et al. (2012) señalan que se realizan pruebas fisicoquímicas y sensoriales al café tostado pero que solo el rendimiento de extracción (proporción o cantidad de sólidos extraídos) y algunas apreciaciones visuales son suficientes para asegurar la calidad. Para Medina & Riaño (2006) una de las acciones más contundentes en la calidad, inicia en un adecuado torrefactado a una temperatura y tiempo determinados donde ocurren muchos cambios físicos y químicos en el grano. Para Cros et al. (1979) es posible caracterizar analíticamente un lote de café a través del estudio únicamente la fracción volátil del grano como índice de calidad del café tostado.

Pero en definitiva, es la evaluación sensorial del café la que permite identificar los defectos presentes en la bebida, medir la intensidad de una característica sensorial como la acidez y el dulzor, y también calificar el sabor, el aroma y la calidad global del producto.

Sin embargo, la aplicación de estos métodos sensoriales puede resultar bastante compleja y costosa. Para los estudios sensoriales del café se requiere primero que todo el personal sea especializado, que cuenten con instalaciones adecuadas, métodos de análisis estandarizados, formularios para el registro de la información en forma escrita o digital, análisis estadísticos y experiencia para la interpretación de los resultados. Con todo ello en norma, se reducen los errores inherentes a los juicios humanos. Por esto, las pruebas sensoriales del café permiten determinar la influencia de los diversos factores y condiciones del proceso del café en la calidad y también permiten conocer las características organolépticas que conforman la calidad en la taza de café.

Pese a todas estas investigaciones, existe muy poca información acerca de estudios previos centrados en la temperatura como uno de los factores que más influye en las características fisicoquímicas del espresso. Es por esto, que sería interesante estudiar el efecto del gradiente de temperatura (perfil de temperatura) en el proceso de obtención del café espresso, ya que es de suponer que los diferentes componentes del café se extraen a diferentes tiempos durante la erogación.

Por otro lado, para la bebida de café especialmente en su modalidad espresso, no se dispone de métodos de control de calidad precisos y de fácil interpretación.

Por ello, si se alcanzara a determinar la influencia de los perfiles de temperatura en las propiedades fisicoquímicas en diferentes tipos de café permitiría potencializar aquellos cafés que no son tan apetecidos en el mercado, ya que estas propiedades estarían asociadas a características organolépticas. Además, se podrían relacionar los datos obtenidos en un panel de valoración sensorial con las determinaciones fisicoquímicas para distintos tipos de café y perfiles de temperatura de extracción lo que permitiría discriminar, correlacionar y/o clasificar muestras de café.

Zona cafetera del valle del cauca

En Colombia, la zona cafetera tiene aproximadamente 936.682 hectáreas de café suave arábico lavado, de las cuales el departamento del Valle del Cauca aporta 69.000 hectáreas distribuidas en los 39 de sus 42 municipios (FNC; 2012-2013). Por su importancia económica y social, el cultivo del café es catalogado como uno de los principales productos agrícolas del país. Para el Ministerio de Agricultura de Colombia y el DANE², el café es la principal fuente de demanda y desarrollo en la región interandina, donde se concentra la mayor parte de la población rural y en la cual se genera cerca de 500.000 empleos directos y alrededor de 5.000 empleos en la parte agroindustrial (Espinal et al. 2005). A parte de los aspectos económicos y sociales está la combinación particular de diversos factores ambientales correspondientes a la agroclimatología del cultivo (suelos, las variedades, el clima, la temperatura, la lluvia) y las prácticas de manejo en su fase de cosecha y poscosecha que finalmente ayudan a determinar la calidad del grano permitiendo expresar en taza, atributos organolépticos específicos. (Espinal et al. 2005).

Por tal motivo, es muy importante determinar un índice de calidad mediante la relación entre las características organolépticas y agronómicas que permita seleccionar e identificar dentro del departamento del Valle del Cauca en Colombia, aquellas zonas cafeteras que se caracterizan por un tener un café de buena calidad.

² Departamento Administrativo Nacional de Estadísticas- DANE. 2010

1.2. CALIDAD DEL CAFÉ

La calidad del café depende de numerosos factores como la especie *Coffea Arábica* (Arábica) o *Coffea Canephora* (Robusta), la variedad, las condiciones edafoclimáticas y el manejo agronómico del cultivo. Sin embargo la calidad del café es el resultado de muchos procesos y operaciones realizados por todas las personas de la cadena del café que efectúan las labores de producción, cosecha y poscosecha.

Cuando se obtiene un café de buena calidad, es porque el café ha presentado características de inocuidad, buena calidad física, buena calidad sensorial y una composición química natural. (Puertas, 2008).

La inocuidad del café garantiza la calidad higiénica tanto de los frutos y los granos como de la bebida, es decir que no deben contener sustancias de origen químico o microbiano en cantidades que puedan causar daño a la salud como por ejemplo, los residuos de plaguicidas, combustibles, mohos, humos y micotoxinas que además de originar defectos en el grano inciden en el aroma y sabor de la bebida.

La calidad física del café se puede determinar mediante la observación de los granos del café pergamino seco (CPS): como la apariencia homogénea, presencia de un color amarillo del pergamino y verde oliva de la almendra, tamaño según las especificaciones de mercado y un contenido de humedad entre el 10 y el 12%. Por el contrario, cuando los granos presentan defectos como guayabas, negros, severamente dañados por la broca, aspecto mohoso, vinagres y, los granos no despulpados o muy decolorados no son aceptables, ya que una vez son procesados causan en la bebida sabores como fermento, *stinker*, fenólico, terroso, mohoso, acre, reposo o contaminado no muy deseables para el consumidor.

La calidad sensorial permite inferir las condiciones bajo las cuales se mantenido el café desde su cultivo hasta la obtención de la bebida. Un café de buena calidad sensorial puede presentar un balance en las características sensoriales (fragancia, aroma, acidez, amargor, cuerpo y sabor) y es imprescindible que no presente olores ni sabores extraños que muestren un deterioro del producto o una contaminación.

Evaluación sensorial del café

Los métodos de la evaluación sensorial se usan básicamente en la industria de los alimentos para determinar la calidad de las materias primas y del producto, en el control de los procesos de fabricación y para los estudios de preferencias en los consumidores. En el café, la evaluación sensorial es el método usado para conocer el aroma, el sabor y la sanidad del café, se denomina también cata o prueba de taza. Por medio de esta técnica se pueden identificar los defectos presentes en la bebida, cuantificar y calificar la intensidad de una característica sensorial.

Las características organolépticas del café se valoran a través de los sentidos **La fragancia** que es la característica del aroma del café tostado y molido. Es la primera cualidad que el catador percibe en la bebida del café al oler. Su intensidad, cualidad y tipo, dan indicios de la calidad del café, de su frescura y de las condiciones de cultivo, así como del proceso para la obtención del producto final. Un buen café presenta un aroma dulce con notas herbales, frutales o a especias. A través del gusto se identifica la acidez, amargor y el dulzor. **La acidez**, es el sabor característico de ciertos ácidos como el acético o de frutas cítricas como el limón o la naranja. Esta es una cualidad propia y muy positiva que presenta la especie *Coffea Arábica L.* cuando se ha beneficiado por la vía húmeda. La acidez resulta indeseable cuando se califica como agria, picante, acre, astringente o ausente y se derivada de inadecuadas prácticas de cosecha y de beneficio del café. **El amargor** es una característica normal del café debida a su composición química, la intensidad depende del grado de torrefacción y de la preparación de la bebida. **El dulzor** es una cualidad propia del café Arábica debida a su composición química.

Cuando se activan los terminales del sentido del tacto, se puede percibir si la bebida está caliente o fría, la astringencia y el cuerpo. **La astringencia** es la sensación de ardor al final de la lengua, la garganta y en la boca del estómago. Básicamente este defecto del café se debe a la gran cantidad de granos inmaduros incluidos en el tostado. **El cuerpo** se refiere a la sensación del café en la boca, a la viscosidad, el peso y el grosor con que es percibido en la lengua. Esa característica se relaciona con los aceites y las sustancias que se extraen durante el tratamiento del grano. Una buena bebida de café presenta cuerpo completo, moderado y balanceado. Por

último se encuentra **la impresión global** que enmarcan todas las propiedades percibidas con el sentido del olfato, el gusto y el tacto. Es la clasificación y calificación general que acepta o rechaza la calidad de un café.

Por ejemplo, cuando la bebida de café es preparada con Robusta presenta sabor muy amargo, aroma fuerte y característico, destacando las nota terrosas, leñosos y cereal, se negocia a menor precio y es utilizado en su mayoría en la preparación de cafés solubles, aunque también es consumido en bebidas solo o en mezclas con otras variedades de café, mientras que los cafés Arábica presentan cualidades suaves, dulces, frutales y características sensoriales y químicas específicas que dependen de la madurez, del tipo de beneficio y de las condiciones procesamiento, almacenamiento y preparación. (Puertas, 2011b; Marín et al. 2003).

Hasta el momento, la evaluación sensorial del café es una herramienta que se ha implementado durante mucho tiempo en el control de calidad del café como la única forma de generar términos descriptivos a través de panelistas entrenados, indispensables en los procesos de comercialización y consumo.

1.3. LA BEBIDA DEL CAFÉ

La bebida de café es un preparado de semillas tostadas y molidas de la planta de café con agua. Está constituido en la parte superior por una capa de pequeñas burbujas de gas dispersas con una emulsión de gotitas de aceite microscópicas y en la parte inferior por una solución acuosa de azúcares, ácidos, proteínas, cafeína y sólidos coloidales.

El café espresso (CE), es una de las mejores formas de consumir el café por que extrae sustancias aromáticas que provoca diferentes tipos de sensaciones organolépticas y que según Andueza et al. (2007) están asociadas a características fisicoquímicas.

A pesar de la popularidad del café espresso en todo el mundo, poco se sabe sobre

el método de preparación y el rol que desempeña el agua y su contenido iónico en la calidad de la bebida, la variedad de café, el beneficio, el punto de molido y su aceptación por parte del consumidor.

El espresso se refiere a la preparación del café elaborado a través de una percolación rápida con una pequeña cantidad de agua caliente (88-95 °C) a una presión determinada de 9 ± 2 bares que pasa a través de una torta comprimida de café tostado y molido finamente. En este tipo de bebidas es muy importante el tiempo de preparación que debe estar entre 25 a 30 segundos y el volumen final que debe ser comprendido entre 25 a 30 mL. Además de estos dos aspectos, el café espresso debe ser cubierto por una gruesa crema marrón aterciopelada de color rojizo llamado crema, un aspecto visual que considera la bebida debidamente preparada y muy apreciada por los conocedores del tema (Illy, 2002).

En cuanto a las características fisicoquímicas del café espresso, estas son responsables de sus propiedades sensoriales peculiares que incluyen un cuerpo fuerte, un buen aroma, un sabor equilibrado entre amargo/ácido y una agradable persistencia, estas características dependen de la selección de café, el molido, el tostado y las condiciones técnicas del proceso de percolación (temperatura, presión y calidad del agua), que debe ser ajustado de acuerdo a tipo café.

1.4 CALIDAD DEL CAFÉ ESPRESSO

Para determinar la calidad de la bebida de café en especial la del espresso, se hace uso también de la evaluación sensorial con ciertos descriptores organolépticos como el cuerpo, la textura de la crema, el olor, el color de la crema, la acidez, el amargo y la astringencia con los cuales se elaboran perfiles de aceptación.

El café espresso (CE) es una de las formas más habituales de consumir el café y su preparación ha sido estudiada por algunos autores, quienes mencionan que la calidad es afectada por diversos factores, entre ellos, **el tostado del café** que para Baggenstoss et al. (2008) es responsable de la sensación organoléptica principal

llamada aroma, ya que durante este proceso se han encontrado 800 compuestos volátiles de los cuales 30 han sido identificados para este atributo. Para Navarini & Rivetti (2010), la calidad del agua ejerce una influencia relevante en el aspecto visual del café principalmente en la crema y en su contenido iónico. La presión para Andueza et al. (2002) conduce a obtener un mayor contenido de odorantes claves relacionados con la frescura y el sabor afrutado, así como la consistencia de la crema en los CE preparados a 9 atm en comparación con los preparados a 7 atm. Finalmente, otros autores han tratado de explicar la calidad de la bebida a través de la determinación de las características fisicoquímicas trabajando a presión y temperatura constantes en todo el tiempo de erogación y han efectuado comparaciones entre el efecto de dichos parámetros en distintas variedades y mezclas de café.

1.4.1 ASPECTOS QUE DETERMINAN LA CALIDAD DE UN ESPRESSO

Calidad del agua

La bebida de café está constituida en un 98% de agua, por lo que por supuesto, su calidad debe afectar críticamente el sabor de la bebida. En el proceso de extracción para obtener una buena bebida, el agua debe ser fresca para conservar todo el aire disuelto en ella, ya que finalmente tanto la composición como el aire son componentes importantes en el sabor y aspecto visual del CE.

El agua también debe comenzar fría en el proceso, porque el agua caliente va perdiendo algo de aire disuelto y puede recoger minerales solubles de los conductos. Además, es de suma importancia utilizar agua del grifo filtrada o agua embotellada lo más neutra posible para que contribuya al sabor del CE y al proceso de extracción. Navarini & Rivetti (2010) referencian que Gardner (1958) encontró que si la concentración de los iones en solución es elevada, se cambia la velocidad con la cual el agua pasa a través de la pasta de café molido comprimida, provocando de este modo una extracción diferencial de los sólidos de café al mismo tiempo que acelera la acumulación de cal perjudicial para la caldera en el interior de la máquina, pero por otro lado una ausencia de minerales dará al CE con un sabor insípido.

Navarini & Rivetti (2010) mencionan que según Pangborn (1971) el agua es sin lugar a duda un ingrediente esencial para la preparación y la calidad del café. Este autor concluye a través de sus ensayos con bebidas de café y té preparado con diferentes calidades de aguas potables, que el contenido de carbonatos de algunas de ellas traía consecuencias poco agradables a la bebida, dando como resultados características insípidas y amargas, pero cuando empleó agua destilada estas bebidas eran excesivamente ácidas y astringentes, lo que más tarde corroboró Sivetz (1972) con estudios sobre la relevancia de la alcalinidad del agua en la influencia (neutralizante) de la acidez del café. Otra investigación realizada por Cammenga & Zielasko (1997) referenciada en el mismo documento, demuestran que la calidad del agua afecta el sabor y que la presencia de sabores extraños debido a los tratamientos de desinfección (cloración) y de la dureza del agua a través de su potencial de escala, afectan indirectamente a la calidad de la bebida mediante la reducción de la eficacia de transferencia de calor de los intercambiadores que afecta a la temperatura de extracción.

Tamaño de partícula - granulometría

En la preparación del CE se busca extraer todos los componentes del café por acción del agua caliente que atraviesa en un tiempo definido de contacto, la pasta de café adecuadamente molida; es así, cuanto mayor sea el tamaño de las partículas que forman la pasta, menor será la superficie de contacto entre el agua y el café por tanto la bebida será muy clara, de igual manera, si se presentan partículas muy finas habrá una mayor superficie de contacto y la bebida puede resultar oscura. (Guevara & castaño, 2005).

Desde el punto de vista fisicoquímico, el tamaño de las partículas en el proceso de preparación de la bebida de café, tiene una gran influencia en las propiedades organolépticas. (Clarke & Macrae, 1987).

Crema

La crema es el atributo más importante y característico que se puede percibir inmediatamente después de la preparación de café espresso. Es una capa de crema en la parte superior de la bebida que está compuesta por dióxido de carbono generado durante el tostado de café y atrapado dentro de la estructura celular (Nunes et al. 1997).

Illy & Navarini (2011) referencian que según Rosa et al. (1986), la crema y su persistencia se ve afectada por el contenido de agua que pasa a través del café tostado molido, el grado de compresión en el soporte del filtro, y la cantidad de café tostado y molido utilizado para su preparación. Por otro lado, Anderson et al. (2003a) muestran que el contenido de dióxido de carbono, medido después de la molienda, varía de 4.0 a 8.6 mg/g de café tostado, con un promedio general de 5,7 mg/g (2,9 mL / g en STP).

Temperatura

Los descriptores sensoriales en un café espresso están influenciados por la temperatura y la presión del agua. Según Albanese et al. (2009) señalan que cuando se modifican estos parámetros, varía la cantidad y el tipo de sustancias que extrae durante el proceso de percolación modifican la presencia y la intensidad las propiedades fisicoquímicas y sensoriales de CE.

1.5. MÉTODOS ESTADÍSTICOS

En cuanto al procedimiento estadístico para determinar la calidad, Duarte et al. (1996) y Riaño (2013) aseguran que se han empleado diversos métodos de análisis estadístico para identificar elementos discriminatorios en la calidad del café. Sin embargo, hay poca información acerca de los métodos estadísticos empleados.

1.5.1 DATOS FALTANTES (MISSING VALUES)

En el análisis de datos es muy común encontrar valores faltantes, así como las observaciones atípicas, esta situación es una limitante para analizarlos mediante métodos estadísticos tradicionales que han sido diseñados para conjuntos de datos completos, ya que afecta sustancialmente la fase de inferencia estadística. Esta falta de datos en los registros obtenidos está asociada a diversas causas: al desconocimiento de la información solicitada, fatiga del informante o rechazo de participar de la persona que diligencia el formato.

Según se cita en Medina & Galván (2007), los valores faltantes forman parte de un conjunto de observaciones con características especiales que incluyen a los datos agrupados, agregados, redondeados censurados o truncados. En estas y otras situaciones se debe diferenciar estos registros para que puedan ser analizados adecuadamente, ya que el uso del análisis normal sin un conocimiento específico, alteraría las medidas de dispersión y de tendencia central (Donner 1982, Little & Rubin 1987, Roth et al. 1996). Aunque existen diferentes tratamientos para los valores faltantes no existe ninguno método que pueda ser recomendado a cabalidad, ya que esto depende de diferentes aspectos como el porcentaje de omisión en la variable, que si es alto, la imputación atentaría contra la confiabilidad estadística del estudio y posteriormente invalidaría las conclusiones del estudio (Medina & Galván, 2007).

1.5.1.1. MECANISMO DE PÉRDIDA DE LOS DATOS FALTANTES

Han surgido muchas metodologías para el tratamiento de la información con datos faltantes dando resultados aceptables. Según Badler et al. (2004), estos datos faltantes en las variables se simulan conforme el proceso en el que se produce la pérdida de la información en un experimento y según las características de la probabilidad de respuesta. Por ello, Rubin (1976), hace una rigurosa definición de los distintos supuestos que se podrían hacer acerca del mecanismo de pérdida de los datos faltantes. Si se supone que hay datos faltantes sobre una determinada variable de Y , se pueden decir que hay tres clases, pueden estar perdidos completamente al azar- MCAR, datos perdidos al azar MAR o no perdidos al azar-

MNAR (Allison, 2002). Todo depende de la información incompleta en la variable, esta se puede presentar de manera aleatoria, o sea, estar ligada a valores correspondientes a otra variable relacionada con la que presenta pérdidas o en categorías de la propia variable, es decir, que los valores no observados sean diferentes a los observados.

La notación para los mecanismos de pérdida es la siguiente, Y (variables dependientes) es una matriz rectangular ($n \times k$), donde n es el número de casos y k es el número de variables, entonces Y_{ij} sería la j – esima variable en el i – esimo caso de la matriz de datos Y , ahora, si hay datos faltantes se puede descomponer la matriz Y en, (Y_{obs}, Y_{mis}, R) , donde Y_{obs} contiene los valores observados de Y , Y_{mis} contiene los valores faltantes de Y , y R está definida como la matriz indicador de respuesta que define las posiciones de valores observados y de los no observados (datos faltantes) en la matriz de datos Y y X son variables independientes (todas observadas):

$$R_{ij} \begin{cases} 0 & \text{si } y_{ij} \text{ es no observado} \\ 1 & \text{si } y_{ij} \text{ es observado} \end{cases} \quad (1)$$

Cualquier mecanismo de pérdida de datos puede definirse mediante la especificación de la distribución condicional del indicador de respuesta dado los datos completos $P(R|y, X)$.

Datos perdidos completamente al azar – MCAR (Missing Completely At Random):

Ocurre cuando la probabilidad de que los valores faltantes en Y no está relacionada con el valor de Y misma o con los valores de otras variables en el conjunto de datos, cuando se cumple esta suposición, el conjunto de individuos con datos completos puede considerarse como una submuestra aleatoria simple del conjunto original de observaciones, dicho de otra forma, si la probabilidad de los datos faltantes es independiente de los datos observados, no observados y de las variables independientes, se dice que el proceso es MCAR, es decir que la ausencia de la información no está originada por ninguna variable presente en la matriz de datos y

R es independiente de y_{obs} y y_{mis} :

$$P(R|y, X) = P(R|X) \forall Y \quad (2)$$

Datos perdidos al azar – MAR (Missing At Random): se dice que los datos en Y son faltantes de forma aleatoria, si la probabilidad de los datos faltantes en Y no está relacionado con el valor de Y después de que se han controlado otras variables en el análisis. Es decir, que los datos perdidos son dependientes o están relacionados únicamente con los datos observados, por lo que su ausencia está asociada a variables presentes en la matriz de datos y R es independiente de y_{mis} :

$$P(R|y, X) = P(R|X, y_{obs}) \forall Y_{mis} \quad (3)$$

Datos no perdidos al azar – MNAR (Missing Not At Random): Ocurre cuando la probabilidad de los datos perdidos en una variable Y depende de los valores de dicha variable una vez que se han controlado el resto de las variables, es decir que la probabilidad de respuesta no es independiente de las variables no observadas completamente y posiblemente tampoco de las observadas:

$$P(R|y) \text{ Depende de } y_{mis} \quad (4)$$

El mecanismo de pérdida (proceso de no respuesta), según Allison (2002) para MCAR y MAR, es ignorable. Esto básicamente significa que no se necesita modelar el mecanismo de datos faltantes, como parte del proceso de estimación y para MNAR es "no ignorable". Tanto MCAR y MAR permiten valores perdidos que dependen de la variable X . Es imposible comprobar si la condición MAR se cumple, ya que dado que no se conocen los valores de los datos que faltan, no se puede comparar los valores de los casos con y sin perder los datos para ver si difieren sistemáticamente en una variable determinada.

1.5.1.2. MÉTODOS PARA EL MANEJO DE DATOS FALTANTES

MÉTODOS CONVENCIONALES

Existen métodos convencionales utilizados para el manejo de datos faltantes, dentro de los cuales podemos mencionar dos, *Listwise* y *Pairwise*, los algoritmos de imputación se describen a continuación.

Listwise (Case Deletion) o eliminación por listas

Es una de las formas más sencillas de obtener un conjunto completo de datos y asume que los datos faltantes siguen un patrón **MCAR**. En este método se elimina sólo los casos con valores perdidos considerándose así el método más fácil y sencillo de hacer frente a los datos faltantes y se basa en la eliminación de los casos incompletos (registros con datos faltantes en una variable) del conjunto de datos. Una ventaja en el uso de eliminación por lista es que todos los análisis se calculan con el mismo conjunto de casos. Pero este método presenta dos grandes inconvenientes. En primer lugar, la pérdida substancial de la información en bases de datos, lo que altera el tamaño de la muestra y el número de variables. (Carter, 2006; Magnani, 2004). El segundo problema es que supone que la observación de los valores que faltan no es importante y puede ser ignorado, o sea que la falta de respuesta se genera de manera aleatoria asumiendo que la submuestra de los datos eliminados tiene la misma distribución que los datos completos lo que en situaciones prácticas no se cumplen y los valores estimados serán sesgados de los parámetros poblaciones lo que podría invalidar las conclusiones. (Carter, 2006; Medina & Galván, 2007).

Pairwise o Eliminación por pareja

Este método incluye todos los datos disponibles y a diferencia del método Listwise, la eliminación por parejas o Pairwise sólo elimina los valores específicos que falta en el análisis (no todo el caso o sujeto). Es decir, todos los datos disponibles se incluyen, contrario al Listwise donde se elimina los casos (sujetos) que tienen valores perdidos en cualquiera de las variables objeto de análisis. Este método

asume un patrón de datos perdidos **MCAR** en los datos omitidos, y hace uso de toda la información disponible sin efectuar ningún tipo de corrección en los factores de expansión. Las observaciones que no tienen datos se eliminan, y se hacen los cálculos con diferentes tamaños de muestra lo que limita la comparación de resultados en este caso, el registro con los valores que faltan sólo se utiliza en los análisis que no requieren de la variable que falta. (Medina & Galván, 2007; Marwala, 2009). Su principal ventaja con respecto al Listwise es que preserva una mayor cantidad de información, aunque a cambio hace complicada la interpretación de los coeficientes de correlación, ya que cada coeficiente tendrá un tamaño muestral diferente. (Varela et al. 1998)

MÉTODO DE IMPUTACIÓN SIMPLE

Los métodos de imputación simple radican en asignar un valor por cada valor faltante, tomando como referencia el valor de la propia variable o de otras variables, generando así, una base de datos completa, la técnica de imputación simple usada en esta aplicación es la **Hot-Deck**.

Hot-Deck o fichero caliente

Es un procedimiento de imputación de datos no paramétrico a través de una distribución no condicionada, que consiste en ubicar registros completos (donantes) e incompletos (receptores) e identificar las características comunes entre donantes y receptores. Consecutivamente se hace la imputación entre observaciones con características comunes, la selección de los donantes se realiza en forma aleatoria, de esta manera la suposición se basa en que dentro de cada grupo de clasificación los valores faltantes siguen la misma distribución como aquellos que no son faltantes. La ventaja de este método es evitar que se introduzcan sesgos en el estimador de la varianza preservando las distribuciones conjuntas y marginales. Cuando los datos faltantes siguen un patrón **MCAR** produce estimaciones consistentes (Ávila, 2002).

Existen otros métodos de imputación simple conocidos como **Medias no**

condicionadas que consiste en imputar cada valor faltante por el promedio de la variable, bajo este procedimiento se asume que los datos faltantes siguen un patrón MCAR, **Medias condicionadas para datos agrupados** que es una variación de las medias no condicionadas, aquí se forman categorías a partir de las covariables correlacionadas con la variable de interés, después se imputan los datos faltantes con observaciones provenientes de una submuestra que comparte características comunes, generando así, tantos promedios como categorías hayan. Asume que los datos faltantes siguen un patrón MCAR, por **Variables ficticias** que consiste en crear una variable indicadora para identificar las observaciones con datos faltante, posteriormente se realizan las estimaciones mediante el modelo de regresión lineal como en Listwise, asignando “0” si no se conoce y “1” de otra manera, por **Regresión** que consiste en eliminar las observaciones con datos incompletos y ajustar una ecuación de regresión para predecir los valores \hat{y} que se usaran para sustituir los valores de Y que faltan, de modo que valor es construido como una media condicionada de las covariables X's y los datos faltantes siguen un patrón MCAR y por **máxima verosimilitud** -MV que asume que los datos faltantes siguen un patrón MAR y se demuestra que la distribución marginal de los registros observados está asociada a una función de verosimilitud para un parámetro desconocido θ , bajo el supuesto de que el modelo es adecuado para el conjunto de datos completo, el procedimiento para estimar los parámetros del modelo utiliza una muestra con datos faltantes y estima los parámetros del modelo con los datos completos empleando la función de máxima verosimilitud, luego utiliza los parámetros estimados para predecir los valores omitidos y posteriormente sustituye los datos por las predicciones obteniendo los nuevos valores de los parámetros maximizando la verosimilitud de la muestra completa hasta lograr que el valor de los parámetros no cambie entre dos iteraciones sucesivas (Useche & Mesa, 2006; Schafer & Graham, 2002; Durrent, 2005; Little y Rubín, 1987).

MÉTODO DE IMPUTACIÓN MÚLTIPLE

Rubín en 1987, introdujo el método de imputación múltiple en donde cada dato faltante debía ser remplazado a partir de $m > 1$ simulaciones a través de un proceso estocástico, aquí se seleccionan posibles valores para la información faltante y se

utilizan dichos valores para obtener un conjunto de datos completo (Alfaro & Fuenzalida, 2009; Durrant, 2005).

Van (2007) menciona que cuando se presentan variables categóricas y continuas, el proceso de imputación múltiple **por ecuaciones encadenas – MICE**, proporciona mejores resultados, en todos los casos la validez del proceso de imputación se basa en el supuesto de que el patrón de distribución de los valores faltantes es aleatorio MCAR o MAR (Mediavilla, 2010). El sustento teórico de la imputación múltiple, se encuentra en la estadística bayesiana, la cual utiliza la información de la muestra para realizar inferencia respecto de los parámetros. Para obtener la inferencia deseada se debe realizar sus estimaciones y combinar los resultados a través de las distintas bases de datos. Para lograr este objetivo, Rubín (1986) propone reglas simples que permiten combinar los resultados obtenidos en las distintas bases imputadas, de modo que sea posible ajustar los errores estándar de los estimadores para considerar la incertidumbre generada de las imputaciones. El método de imputación múltiple no soluciona el problema de la información faltante, sino que lo acomoda desde una perspectiva estadística, es decir que se podrá contar con información completa, pero se deberá manejar múltiples bases de datos donde cada una de ellas tiene un valor posible para la observación faltante. El investigador entonces deberá desarrollar su análisis en cada una de las m bases de datos completas y luego combinar los resultados a fin de obtener las conclusiones finales de su investigación. Según Alfaro & Fuenzalida (2009), mencionan que el número óptimo de bases de datos (m) depende del porcentaje de información faltante, pero según Allison (2001), en la práctica la elección comúnmente utilizada es $m = 5$, sin embargo, StataCorp (2009) establece que $m = 20$ es un número razonable de imputaciones ya que minimiza de forma empírica el rango de incertidumbre de la estimación del intervalo de confianza.

Descripción De La Técnica De Imputación Múltiple - MI

La imputación múltiple es un método de Monte Carlo descrito por Rubín (1987) con el contexto de emplearla en la falta de respuesta en las encuestas por muestreo, la técnica es muy general y puede ser fácilmente utilizada y que conducen a una única

instrucción de inferencia sobre los parámetros de interés, el procedimiento de imputación múltiple se describe a continuación.

Q es la cantidad escalar de interés en el estudio para ser estimada, por ejemplo, la media población de Y , coeficiente de regresión, la correlación, etc. Con m imputaciones, se puede calcular m diferentes versiones de \hat{Q} y U , sea:

$$\hat{Q}^t = \hat{Q}(Y_{obs}, Y_{mis}^t) \text{ y } U^t = U(Y_{obs}, Y_{mis}^t) \quad (5)$$

Donde \hat{Q}^t es la estimación estadística de Q y U^t su varianza estimada utilizando el t -ésimo conjunto de datos imputados $t = 1, 2, \dots, m$.

La estimación puntual de las imputaciones para Q es simplemente el promedio de las estimaciones puntuales de los datos completos:

$$\bar{Q} = \frac{1}{m} \sum_{t=1}^m \hat{Q}^t \quad (6)$$

Para obtener el error estándar de \bar{Q} uno debe calcular la variación dentro de las imputaciones y entre las imputaciones:

La variación dentro de las imputaciones, que es el promedio de las estimaciones de la varianza de los datos completos:

$$\bar{U} = \frac{1}{m} \sum_{t=1}^m U^t \quad (7)$$

La variación entre las imputaciones que es la varianza de las estimaciones puntuales de los datos completos,

$$B = \frac{1}{m-1} \sum_{t=1}^m (\hat{Q}^t - \bar{Q})^2 \quad (8)$$

La varianza total es definida como:

$$T = \bar{U} + (1 + m^{-1})B \quad (9)$$

El intervalo estimado y el nivel de significancia para el escalar Q son formados

usando una distribución de referencia t y las inferencias son basadas en la aproximación:

$$T^{-\frac{1}{2}}(Q - \bar{Q}) \sim t_v \quad (10)$$

Donde los grados de libertad son dados por:

$$v = (m - 1) \left[1 + \frac{\bar{U}}{(1+m^{-1})B} \right]^2 \quad (11)$$

Así, el intervalo de estimación $100(1 - \alpha)\%$ para Q es:

$$\bar{Q} \pm t_{v, 1-\alpha/2} \sqrt{T} \quad (11)$$

y un p-valor para probar la hipótesis nula $\bar{Q} = Q'$ en contra de una alternativa bilateral es:

$$2P(t_v \geq |\bar{Q} - Q'| T^{-\frac{1}{2}}) \text{ o equivalentemente, } P \left[F_{1,v} \geq T^{-\frac{1}{2}} (\bar{Q} - Q')^2 \right]$$

Ahora, cuando se presentan datos faltantes se tiene en cuenta lo siguiente:

Nótese que los grados de libertad de la ecuación (6) Dependen no sólo de m , sino también de la relación:

$$r = \frac{(1+m^{-1})B}{\bar{U}} \quad (12)$$

Rubín llama a r el *incremento relativo de la varianza debido a la falta de información* o más precisamente la varianza condicional dado $B_\infty = B$, donde \bar{U} representa la estimación total de la varianza cuando no hay información faltante en Q (por ejemplo cuando $B = 0$). Cuando m es grande y/o r es pequeño, los grados de libertad son grandes y se aproxima a una normal. Sin información faltante, la parte posterior llegar a ser normal con media \bar{Q} y varianza \bar{U} , para lo cual la información es \bar{U}^{-1} .

Resultando en:

$$\begin{aligned} \hat{\lambda} &= (\bar{U}^{-1} (v + 1)(v + 3)^{-1} T^{-1}) \bar{U} \quad (13) \\ &= \frac{r + 2/(v + 3)}{r + 1} \end{aligned}$$

Que es una estimación de la fracción de información faltante sobre Q . En las aplicaciones, el cálculo de r y $\hat{\lambda}$ es muy recomendable, ya que son interesantes y útiles en el diagnóstico para evaluar la forma en que los datos faltantes contribuyen a la incertidumbre inferencial de Q .

Finalmente, Rubín & Schenker (1986) demuestran que este método conduce a inferencias que están bien calibrados desde un punto de vista frecuentista y que las estimaciones del intervalo de las imputaciones múltiples cubren el parámetro real al menos el 95% del tiempo, incluso cuando se tiene un m tan pequeño como 2. La razón fundamental para esto en primer lugar, es que la imputación múltiple se basa en la simulación para resolver problema del aspecto de los datos perdidos. Como con cualquier método de simulación, una forma efectiva podría eliminar el error Monte Carlo eligiendo un m lo suficientemente grande, pero, con la imputación múltiple la ganancia resultante de la eficiencia típicamente sería poco importante debido a que el error de Monte Carlo es una parte relativamente pequeña de la incertidumbre inferencial general y la otra razón, son las reglas para combinar los m análisis de datos completos para el conteo de forma explícitamente del error Monte Carlo. Ahora, cuando hay muestras grandes como se demuestra en Rubín 1976, los procedimientos de imputación conducen a inferencias que son válidas porque la inferencia en los datos completos se basa en la aleatorización cuando hay ausencia de falta de respuesta y además, porque el método de imputación es correcto. Ahora, se puede calcular la eficiencia relativa cuando se utiliza un número finito de imputaciones adecuadas m , en lugar de un número infinito.

$$RE = (1 + \lambda / m)^{-1/2} \quad (14)$$

Donde el escalar λ , es la fracción de la información faltante y la eficiencia relativa- RE de un punto estimado es dada unidades de errores estándar. Cuando $\lambda = 0,2$, por ejemplo, una estimación basada en $m = 3$ imputaciones tenderá a tener un error estándar sólo en un $\sqrt{1 + (0.2 / 3)} = 1.033$ veces más grande que el estimado con $m = \infty$ ó con $\lambda = 0.5$, una estimación basada en $m = 5$ imputaciones tienden a

tener un error estándar de $\sqrt{1 + (0.5 / 5)} = 1.0491$ veces más grande que el estimado con $m = \infty$. Por ello las imputaciones con m pequeño son tal validas como las obtenidas con un m grande.

Imputación Múltiple por Ecuaciones Encadenadas – MICE

Según la publicación de Azur, et al. (2011) y Van & Groothuis (2011), las ecuaciones encadenados se utilizan para imputar valores faltantes cuando las variables pueden ser de diferentes tipos y los patrones de valores perdidos son arbitrarios. Para hacer frente a los problemas planteados por la complejidad real de los datos, la especificación se produce al nivel variable. Hipotéticamente los datos completos Y son muestras aleatorias parcialmente observadas de la P-Variable de la distribución multivariada $P(Y_{obs}|\theta)$, se supondrá entonces que la distribución multivariante de Y está completamente especificada por θ , explícita o implícitamente. Las ecuaciones encadenadas proponen para obtener una distribución posterior de θ , mediante el muestreo de forma iterativa a partir de distribuciones condicionales de la forma:

$$P(Y_1|Y_{-1}, \theta_1) , P(Y_2|Y_{-2}, \theta_2) , \dots\dots\dots P(Y_p|Y_{-p}, \theta_p) \quad (15)$$

Donde $Y = (Y_1, Y_2, \dots, Y_p)$ es un conjunto de p variables aleatorias, Y_j con $(j = 1, 2, \dots, p)$ es una de p variables incompletas y $Y_{-j}=(Y_1, Y_2, \dots Y_{j-1}, Y_{j+1}, \dots Y_p)$ es la colección de $p - 1$ variables en Y excepto Y_j , los parámetros $\theta_1, \theta_2, \theta_3 \dots, \theta_p$ son específicos de las densidades condicionales respectivas y no necesariamente son el producto de una factorización de la "verdadera" distribución conjunta $P(Y_{obs}|\theta)$.

A partir de una simple extracción de la distribución marginal observada, la $t - esima$ iteración de la cadena de las ecuaciones se realiza mediante el muestreador de Gibbs:

$$\begin{aligned} \theta_1^{*t} &\sim P\left(\theta_1 \mid Y_1^{(obs)}, Y_2^{(t-1)}, \dots, Y_p^{(t-1)}\right) \\ Y_1^{*t} &\sim P\left(Y_1 \mid Y_1^{(obs)}, Y_2^{(t-1)}, \dots, Y_p^{(t-1)}, \theta_1^{*t}\right) \end{aligned} \quad (16)$$

$$\theta_p^{*t} \sim P(\theta_p \mid Y_p^{(obs)}, Y_1^{(t)}, Y_2^{(t)}, \dots, Y_{p-1}^{(t)})$$

$$Y_p^{*t} \sim P(Y_p \mid Y_p^{(obs)}, Y_1^{(t)}, Y_2^{(t)}, \dots, Y_{p-1}^{(t)}, \theta_p^{*t})$$

donde $Y_j^{(t)} = (Y_j^{(obs)}, Y_j^{*(t)})$ es la j -ésima variable imputada para la iteración t , se puede observar que las imputaciones anteriores $Y_j^{*(t-1)}$ sólo entran $Y_j^{*(t)}$ a través de su relación con otras variables, y no directamente. El nombre de *ecuaciones encadenadas* se refiere al hecho de que el muestreador de Gibbs se puede implementar fácilmente como una concatenación de los procedimientos univariantes para completar los datos faltantes.

Para Raghunathan et al. (2001), MICE es entonces, una técnica de imputación múltiple, que opera bajo el supuesto de que, dadas las variables utilizadas en el procedimiento de imputación, los datos faltantes están ausentes de forma aleatoria (MAR), lo que significa que la probabilidad de que un valor no esté presente sólo depende de los valores observados y no de los valores no observados. En otras palabras, después de controlar todos los datos disponibles (es decir, las variables incluidas en el modelo de imputación) "cualquier valor faltante restante es completamente al azar" y cuando los datos no son perdidos MAR podría dar lugar a estimaciones sesgadas (Schafer & Graham, 2002; Graham, 2009; Stuart et al. 2009).

Pasos Imputación Múltiple por Ecuaciones Encadenadas- MICE

En el proceso de ecuación en cadena el problema es extraer las imputación de $P(Y)$, la distribución incondicional multivariado de Y . Sea t un contador de iteración. Se puede repetir la siguiente secuencia de iteraciones utilizando el muestreador de Gibbs:

Para Y_1 : extraer imputaciones de Y_1^{t+1} de $P(Y_1 \mid Y_2^t, Y_3^t, \dots, Y_p^t)$

Para Y_2 : extraer imputaciones de Y_2^{t+1} de $P(Y_2 \mid Y_1^{t+1}, Y_3^t, \dots, Y_p^t)$

Para Y_j : extraer imputaciones de Y_j^{t+1} de $P(Y_j \mid Y_1^{t+1}, Y_2^{t+1}, \dots, Y_{j-1}^t)$

Con las ecuaciones (16) se puede deducir los siguientes pasos:

(1) Se realiza una imputación simple, como la imputación de la media, para cada valor faltante en el conjunto de datos. (2) Estas medias imputadas pueden ser consideradas como "marcadores de posición". El "marcador de posición" imputación de la media para una variable (Y_j) es un conjunto de regresos para los valores faltantes. (3) Los valores observados de la variable " Y_j " en (2) son regresores en las otras variables en el modelo de imputación, que pueden o no pueden consistir en todas las variables del conjunto de datos. En otras palabras, " Y_j " es la variable dependiente en un modelo de regresión y todas las otras variables son variables independientes en dicho modelo. Estos modelos de regresión operan bajo los mismos supuestos que uno hace cuando se realizan los modelos lineales de regresión logística, o Poisson, es decir, fuera del contexto de imputación de datos faltantes. (4) Los valores faltantes para " Y_j " se reemplazan con las predicciones (imputaciones) del modelo de regresión. Cuando se utiliza " Y_j " posteriormente como una variable independiente en los modelos de regresión de las variables, tanto los valores observados como los imputados son utilizados. (5) Los pasos (2-4) se repiten para cada variable con datos faltantes. El ciclaje a través de cada una de las variables constituye una iteración o "ciclo". Al final de un ciclo de todos los valores faltante se han sustituido con las predicciones de las regresiones que reflejan las relaciones observadas en los datos. Estos mismos pasos se repiten para un número de ciclos en donde en cada ciclo las imputaciones son actualizadas y alimentadas.

Al final de estos ciclos las imputaciones finales se conservan, dando como resultando un conjunto de datos completo. La idea es que al final de los ciclos la distribución de los parámetros que rigen las imputaciones haya convergido en el sentido de convertirse en estable. Esto, porque de cierta manera, evita la dependencia del orden en el que las variables se imputan. Una vez que el número designado de ciclos se ha completado, el proceso de imputación completo se repite para generar múltiples conjuntos de datos imputados. Los datos observados, evidentemente, serán los mismos a través de los conjuntos de datos imputados; sólo los valores que originalmente habían sido faltantes serán diferentes.

Ahora, una correcta especificación del modelo, puede mejorar aún más el modelo de imputación y la creación de imputaciones válidas, para ello, se deben identificar las condiciones bajo las cuales una variable debe o no debe ser imputada. Esto podría ocurrir si hay falta de diseño en los datos o si hay cuestiones jerárquicas (También llamado salto de patrones).

Después de la especificación del modelo, este es revisado y perfeccionando para posteriormente evaluar el procedimiento de imputación. Esto puede hacerse a través del Resumen de estadísticas (summary) que proporcionan información sobre los valores observados, los valores imputados donde se pueden comparan las diferencias de medias y desviaciones estándar, medidas muy útiles para la identificación de las variables problemáticas y las variables que hay que modificar antes de su uso en los análisis. Otras comparaciones gráficas y numéricas entre los datos observados e imputados que ofrecen representaciones visuales de la medida en que los valores imputados se diferencian de los valores observados y que son muy útiles en el diagnóstico de problemas (Abayomi et al. 2008), son los histogramas, los diagramas de cuantil-cuantil y los gráficos de densidad.

Finalmente, según Stuart et al. (2009) es importante reconocer que variables presentan diferencias muy marcadas entre los valores observados y los imputados, pero esta magnitud no necesariamente significa que las imputaciones no son exactas, sino que de hecho puede ser razonables debido a que puede ser indicativo del sesgo de las imputaciones que se están tratando de resolver.

1.5.1.3 DISCUSIÓN DE LOS MÉTODOS DE IMPUTACIÓN

Para seleccionar un buen método de imputación se debe tener claro que algunas técnicas podrían dar mejores aproximaciones a los valores reales de la investigación, por eso, en la Tabla 1, se muestra las ventajas y desventajas de los métodos de imputación (Listwise o case deletion- CD, Hot Deck -HD e Imputación Múltiple – MI) seleccionados para este estudio. Sin embargo, no hay reglas concretas, pero esto dependerá del tipo y tamaño del conjunto de datos, de los objetivos de la investigación, características específicas de la población,

características generales de cada variable, etc. Goicoechea (2002), indica los criterios que se deben tener en cuenta para seleccionar una técnica de imputación adecuada y se describen a continuación.

Determinar el tipo de variable a imputar: Si es continua, tomar en cuenta el Intervalo para la cual se define, y si es cualitativa, tanto nominal como ordinal, las categorías de las variables.

Seleccionar los parámetros que se desean estimar: Si se desea conocer sólo valores agregados como la media y el total, se pueden aplicar métodos sencillos como imputación con la media o moda, sin embargo, puede haber subestimación de la varianza. En caso de que se requiera mantener la distribución de frecuencia de la variable y las asociaciones entre las distintas variables, se deben emplear métodos más elaborados aplicando imputación de todas las variables faltantes del registro.

Calcular las tasas de no respuesta y exactitud necesaria: Cuando el porcentaje de no respuesta es alto en una base de datos, se considera que no hay confiabilidad en los resultados que se obtengan con el análisis de esta base.

Información auxiliar disponible: Con esta información se puede deducir información de los valores ausentes de una variable o hallar grupos homogéneos respecto a una variable auxiliar que se encuentre altamente correlacionada con la variable a imputar, y de esta manera encontrar un donante adecuado que sea similar al registro receptor.

Para evaluar el método de imputación el mismo autor sugiere los siguientes pasos:

- Una vez que se dispone de un archivo con datos faltantes, se recopila y valida toda la información auxiliar disponible que pueda ser de ayuda para la imputación.
- Se estudia el patrón de pérdida de respuesta.
- Posteriormente se observa si hay un gran número de registros que simultáneamente tienen no respuesta en un conjunto de variables.
- Se seleccionan varios métodos de imputación posibles y se contrastan los

resultados.

- Se calculan las varianzas para los distintos métodos de imputación seleccionados con el objetivo de obtener estimaciones con el mínimo sesgo y la mejor precisión.
- Se concluye a partir de los resultados obtenidos.

Cabe anotar que la validez del método se sustenta en la forma como se obtienen las imputaciones, ya que los estimadores finales se calculan como un promedio de los valores sustituidos lo cual garantiza que se mantenga la variabilidad en los valores imputados.

Tabla 1. Ventajas y desventajas de los métodos de imputación CD, HD y MI.
(Según autores como Otero (2011); Goicochea (2002); Amón (2010))

Listwise: Eliminación de registros con valores faltantes.

Hot - deck: Es un procedimiento de duplicación. Cuando falta información en un registro se duplica un valor ya existente en la muestra para reemplazarlo.

Imputación Múltiple: se basa en la imputación de más de un valor para cada valor ausente. MI consiste en generar $m > 1$ valores aleatorios para cada valor perdido por no respuesta de manera que se dispone de m conjuntos de datos completos

Ventajas	Desventajas	Ventajas	Desventajas	Ventajas	Desventajas
Facilidad de su implementación y posibilidad de comparar los estadísticos univariante	Perdida de información, asume que la submuestra de datos excluidos tiene las mismas características que los datos completos	Los procedimientos conducen a una post-estratificación sencilla. No presentan problemas a la hora de encajar conjuntos de datos No se necesitan supuestos fuertes para estimar los valores individuales de las respuestas que faltan. Conserva la distribución de la Variable.	Cuando hay muchos datos faltantes se reporta muchas veces mismo valor duplicado. Distorsiona la relación con el resto de las variables Carece de un mecanismo de probabilidad Requiere tomar decisiones subjetivas que afectan a la calidad de los datos, lo que imposibilita calcular su confianza Proporciona estimaciones sesgadas de la media.	Incrementa la eficiencia de los estimadores ya que minimiza los errores estándares Obtiene inferencias validas simplemente mediante la combinación de las inferencias conseguidas con las bases de datos completas. Estudia directamente la sensibilidad de las inferencias usando los métodos de las bases de datos completas.	Requiere un mayor esfuerzo y tiempo para crear las bases de datos imputadas. No produce una única respuesta, el investigador deberá manejar múltiples bases de datos donde cada una de ellas tiene un valor posible para la observación faltante.

1.6. ANÁLISIS FACTORIAL

Es una técnica estadística multivariada que tiene como principal objetivo sintetizar y explicar las interrelaciones observadas entre un conjunto de variables observadas mediante un pequeño número de variables latentes no observadas (no medidas). Estas variables aleatorias inobservables, llamadas **factores comunes**, se emplean para explicar todas las covarianzas o correlaciones y cualquier porción de la varianza que es inexplicada por dichos factores se asigna a términos de errores residuales llamados factores **únicos o específicos**. Estos factores no deben estar correlacionados entre sí y es posible que la búsqueda de ellos se confunda con la búsqueda de las componentes principales de las variables, sin embargo, el análisis de componentes principales (ACP) se construyen para explicar las varianzas entre las variables, mientras que, los factores se construyen para explicar las covarianzas o correlaciones entre las variables, también el ACP es una herramienta descriptiva o de tipo exploratoria, mientras que el análisis factorial (AF) exige un modelo estadístico formal para la generación de datos, y por lo tanto acude a las pruebas de hipótesis y a la inferencia. El AF puede ser un análisis exploratorio cuando no se conocen a priori el número de factores y es en la aplicación empírica donde se determina este número y puede ser de tipo confirmatorio cuando los factores están fijados a priori, utilizándose contrastes de hipótesis para su corroboración. (Pérez & Santín, 2007; Kline, 2011).

Los supuestos que fundamentan la técnica implican, primero que factores pueden usarse para explicar fenómenos complejos y segundo, que las correlaciones observadas entre las variables son el resultado del hecho que tales variables comparten los mismos factores. Por lo tanto, caben dos enfoques en el análisis factorial, el primero, cuando se analiza toda la varianza, la común o compartida y la no común (Varianza específica de cada variable + Varianza debida a errores de medición), en este caso el método más empleado, es el de análisis de **Componentes Principales – ACP**. En el segundo enfoque, se puede analizar solamente la varianza común o compartida, en este caso se sustituyen los 1 de la diagonal de la matriz correlaciones por las estimaciones de la varianza que cada ítem tiene en común con los demás y que se denominan comunalidades. Para estimar las comunalidades no hay un cálculo definido y para ello existen diversos

procedimientos como las correlaciones múltiples o coeficientes de fiabilidad. A este procedimiento se le denominan análisis de factores comunes – AFC. Los pasos en el AF según Grajales, 2000; Morales, 2013; Morales & Martínez, 2008, comienzan con:

Análisis de la matriz de correlación

Se computa una matriz de correlación de todas las variables con el objetivo de identificar las variables no correlacionadas con otras. Las variables que tienen pequeñas o nulas correlaciones entre sí, son aquellas que no comparten factores en común. Por medio de la **prueba de esfericidad de Bartlett** se prueba la hipótesis de que la matriz de correlación es una matriz de identidad (*todos los valores en la diagonal son 1 y todos fuera de la diagonal son 0*). Es decir que con esta prueba, se muestra la probabilidad estadística de que la matriz de correlación tiene correlaciones significativas al menos entre algunas variables. La significancia debe ser menor a 0.05 para proceder con la técnica. Otra forma de observar el grado de correlación entre las variables es por medio de los **coeficientes de correlación parcial** que son estimados de la correlación entre los factores únicos y deben ser cercanos a cero. Cuando se cumplen las suposiciones del análisis factorial, la matriz que contiene los coeficientes de correlación parciales negativos (*La matriz de correlaciones anti-imágen*) debe mostrar una proporción muy reducida de coeficientes de correlación altos, a fin de que pueda considerarse el análisis factorial. Además, existe una **medida de la adecuación de la muestra (MSA)** que a través de un índice compara las magnitudes de los coeficientes de correlación observados y las magnitudes de los coeficientes de la correlación parcial, es decir que cuantifica el grado de intercorrelación entre las variables y lo apropiado del análisis de factor. Si una variable tiene una MSA igual 1, puede ser predicha de manera perfecta sin error por las otras variables en el estudio. Otra prueba de adecuación de la muestra se denomina **Káiser-Meyer-Olkin (KMO)** en la cual su índice nos indica que cuando presenta valores pequeños (cercanos a cero), no es recomendable usar AF, ya que las correlaciones entre pares de variables no son explicadas por otras variables. El **coeficiente R^2** o coeficiente de correlación múltiple al cuadrado entre una variable y el resto, es otro indicador de la fortaleza de la asociación lineal entre las variables y es reconocido como Comunalidad. Cuando este coeficiente es pequeño para un

variable en particular, es recomendable considerar la posibilidad de eliminarla del conjunto de variables en estudio.

Extracción factorial

El objetivo es establecer los factores que subyacen en las variables medidas u observadas y existen varias técnicas que se emplean, como el ACP donde se forma una combinación lineal de las variables observadas con el propósito de seleccionar el número de factores necesarios para representar los datos y que almacena el porcentaje total de la varianza que es explicada por cada uno de ellos, aquí la primera componente principal recoge la mayor cantidad de la varianza muestral total, a la segunda componente principal le corresponde la siguiente cantidad de varianza inmediatamente inferior a la primera y no está correlacionada con la primera componente. Así sucesivamente el resto de las componentes explican proporciones menores de la varianza muestral total, cabe aclarar que la varianza total es la suma de las varianzas de cada variable. El **AFC** analiza la varianza común o compartida de todas las variables en factores, donde cada factor está compuesto por todos los ítems, pero en cada instrumento los ítems tienen un peso específico distinto según sea su relación con cada factor. La varianza de todas estas nuevas medidas es equivalente a la máxima varianza de la medida original que es posible explicar. Por ello se dice que en el AFC se reduce a la búsqueda de los pesos específicos atribuidos a cada factor. Los factores con coeficiente grandes (en valores absolutos) para una variable son factores que están estrechamente relacionados con dicha variable. Esta información se puede observar en la *matriz de estructura factorial* que muestra las correlaciones simples entre las variables y los factores además de contener las correlaciones entre los factores, pero estos pesos contienen la varianza única o compartida entre las variables y los factores más las correlaciones entre los factores, cuando estas correlaciones entre los factores tienden a ser más grandes, se hace más difícil distinguir cuáles son variables que aportan un mayor peso en cada factor.

Cuando los factores estimados no se han correlacionado entre sí, es decir, son ortogonales, se puede calcular la proporción de la varianza de cada variable que es

explicada por el modelo factorial para emitir un juicio sobre la bondad del modelo factorial al describir las variables originales. Esto porque al no estar correlacionados los factores, el total de la proporción de varianza explicada es simplemente la suma de las proporciones de las varianzas explicadas por cada factor. Finalmente, el objetivo de la fase de extracción de los factores es determinar el número de factores comunes necesarios para describir los datos y es una decisión que se basa en los valores propios (eigenvalues) y el porcentaje de la varianza total que aporta cada uno de los diferentes factores.

La fase de rotación

Con la rotación de los factores se busca transformar la matriz inicial en otra matriz que sea más fácil de interpretar para poder identificar los factores que son substancialmente significativos. Existen dos tipos de rotación la ortogonal o la oblicua. La **rotación es ortogonal** cuando los ejes de coordenadas se giran manteniendo un ángulo de 90 grados entre ellos y esto asume que los factores que son identificados no se relacionan entre sí, mientras que una **rotación oblicua** que ocurre cuando los ejes que se rotan conservan entre sí un ángulo diferente a 90 grados y se asume que hay cierto grado de relación entre los factores que lleguen a conformarse. La rotación de los ejes no afecta la bondad de la solución factorial, y aunque la matriz factorial cambia así como los porcentajes atribuibles a cada factor, las comunalidades y los porcentajes de la varianza total explicada no varían, es decir, la rotación solo redistribuye la varianza explicada por los factores individuales. La forma de calcular la matriz factorial da lugar a los distintos métodos de rotación ortogonales y oblicuos de los cuáles los más utilizados son los siguientes: el **Método Varimax** que minimiza el número de variables que tienen una alta carga en un factor para fortalecer la interpretación de dichos los factores. Este método busca aumentar varianza de las cargas factoriales al cuadrado de cada factor logrando que algunas de sus cargas factoriales tiendan a aproximarse a uno, mientras que otras se acerquen a cero, logrando un sentido de pertenencia más palpable de cada variable a ese factor, el **Método Quartimax** minimiza el factores necesarios para explicar la variable, donde cada variable tiene una alta correlación con un pequeño número de factores a través de la maximización de la varianza de las cargas factoriales al

cuadrado de cada uno de los factores, es decir, que el método logra que cada variable concentre su pertenencia en un factor determinado a través de una carga factorial alta mientras que en el resto de los factores, las cargas factoriales tiendan a ser bajas. En cuanto a la interpretación con el método se suele ser más claro en relación a la comunalidad total de cada variable, el **Método Equamax** es una combinación de los dos anteriores, que busca maximizar la media de los criterios anteriores a través de la simplificación de los factores y las variables. Para su interpretación es necesario agrupar las variables que tienen una alta carga respecto al mismo factor o también se puede requerir que las cargas pequeñas sean omitidas, posteriormente es indispensable un esfuerzo teórico lógico por parte del investigador para encontrar y demostrar el significado y el sentido de los resultados, el **Método Promax** consiste en alterar los resultados obtenidos de una rotación ortogonal hasta poder crear una solución con cargas factoriales que sean lo más próximas a la estructura ideal y se obtiene elevando las cargas factoriales obtenidas a una potencia entre 2 y 4 en una rotación ortogonal, esto porque entre mayor sea la potencia más oblicua será la solución obtenida y el **Método Oblimin directo** que es un método de rotación oblicua no ortogonal. En esta rotación, las cargas factoriales no coinciden con las correlaciones entre el factores y las variables (matriz de configuración) por lo que se apoya en un valor delta que permite controlar el grado de oblicuidad que pueden alcanzar los factores. Cuando un valor de delta tiende a ser más negativo, los factores son menos oblicuos, ya que los factores están correlacionados entre sí. Por eso cuando las correlaciones entre los factores son muy altas en la matriz de correlaciones, es conveniente emplear las rotaciones oblicuas, como este método que minimiza la suma productos de los cuadrados de los coeficientes de la matriz de configuración.

Cálculo de puntuaciones factoriales

Después de rotar los factores se calcula la *matriz de puntuaciones factoriales* con el fin de conocer, **a)** los sujetos más extremos o atípicos a través de la representación gráfica de las puntuaciones factoriales para cada par de ejes factoriales. **b)** la ubicación de ciertos grupos en la muestra. **c)** en qué factores sobresalen o no algunos sujetos y **d)** la interpretación, análisis o explicación de la información

obtenida, y de los valores que toman los factores en cada observación. Existen muchos métodos para estimar la *matriz de puntuaciones factoriales* pero los factores estimados deben cumplir algunas propiedades que son deseables, como que cada factor estimado tenga una alta correlación con el verdadero factor, que tengan una nula correlación con los demás factores verdaderos, que estén incorrelacionados o mutuamente ortogonales si son ortogonales y que sean estimadores insesgados de los verdaderos factores. No obstante, el problema de la estimación de los factores comunes es complejo y en la mayoría de los escenarios, no hay una solución exacta ni única. Los métodos de estimación de las puntuaciones factoriales más empleados son el **método de regresión** que estima la *matriz de puntuaciones* mediante el método de los mínimos cuadrados y da lugar a puntuaciones con una máxima correlación con las puntuaciones teóricas. Pero el estimador no es insesgado, ni unívoco y, en el caso de que los factores sean ortogonales, puede dar lugar a puntuaciones correladas, el **método de Barlett** que utiliza el método de los mínimos cuadrados generalizados para estimar dicha matriz que genera puntuaciones correladas con las puntuaciones teóricas, insesgadas y unívocas, pero cuando los factores son ortogonales, se da lugar a puntuaciones correladas y el **método de Anderson-Rubín** que estima la *matriz de puntuaciones factoriales* mediante el método de los mínimos cuadrados generalizados imponiendo la condición adicional en donde las puntuaciones ortogonales obtenidas están correladas con las puntuaciones teóricas, pero el estimador no es insesgado ni es unívoco.

1.6.1. ANÁLISIS FACTORIAL BAYESIANO

El análisis Bayesiano combina las distribuciones a priori de los parámetros con la probabilidad de los datos para formar distribuciones a posteriori de los parámetros estimados. Las a priori puede ser difusas (no informativa) o informativas, donde la información debe provenir de estudios previos. La parte posterior brinda una estimación en la forma de una media, la mediana o la moda de la distribución posterior. El enfoque Bayesiano también puede utilizar las cadenas de Markov con algoritmo Montecarlo - MCMC para extraer y obtener una distribución marginal posterior de cada cantidad estimada. La idea detrás del uso de las MCMC es que la distribución condicional de un conjunto de parámetros pueda ser usada para extraer

aleatoriamente valores de parámetros, finalmente el resultando es una aproximación de la distribución conjunta de todos los parámetros (Silva & Benavides, 2001; Quinn, 2004).

Hay cuatro puntos clave que motivan a realizar un análisis Bayesiano y son los siguientes: Se puede aprender sobre las estimaciones de parámetros y modelos de ajuste que se ilustra mediante estimaciones de los parámetros que no tienen una distribución normal. El mejor rendimiento se puede obtener en muestras pequeñas, es decir que las muestras grandes teóricamente no son necesarias. Los análisis pueden ser menos exigentes computacionalmente y se obtienen estimaciones altamente confiables. Los nuevos tipos de modelos pueden ser analizados, ya que se pueden estudiar modelos con un gran número de parámetros o donde la máxima verosimilitud no proporcione una aproximación natural.

Este tipo de modelo de análisis factorial es usado cuando se quiere identificar el número y naturaleza de los factores fundamentales responsables de las covariaciones en las p variables observables. Según Quinn (2004), en muchos casos, algunos las puntuaciones observadas de las variables son continuas, mientras que en otras son ordinales, cuando se presenta estas situaciones, los investigadores a menudo ven dentro de sus opciones de análisis de datos las siguientes apreciaciones: **(a)** hacer un tratamiento de las variables ordinales como continuas y usar un modelo de análisis factorial normal con teoría estándar, **(b)** discretizar las variables continuas y usar un modelo de ítem respuesta, **(c)** descartar las variables ordinales (continuas) observadas y trabajar sólo con variables continuas (ordinales), **(d)** renunciar a una estrategia de medición basada en el modelo completo. Pero cada uno de estos movimientos resulta en consecuencias negativas (posiblemente graves) para la correcta inferencia. Cuando se hace caso omiso de la naturaleza ordinal de algunas de las variables observadas se puede dar lugar a estimaciones falsamente precisas y posiblemente sesgadas. Con la discretización de las variables continuas, se deshace la información y se disminuye la precisión de las estimaciones. Descartar algunas variables de manera similar, reduce la precisión de las estimaciones y puede hacer que sea imposible de descubrir la estructura compleja y multidimensional de los datos observados, por

último, los métodos que no basados en los modelos de medición no pueden dar cuenta de los errores de medición y no permitirán a los investigadores evaluar la incertidumbre en las medidas resultantes.

Ante la situación anterior, Quinn (2004), se centra en estos problemas mediante la formulación de un modelo de medición que sea apropiado para respuestas multivariantes que tienen algunas componentes continuas, algunas ordinales o para combinaciones continuas y ordinales, permitiendo una sencilla interpretación de los parámetros del modelo. El autor toma un enfoque bayesiano para este modelo formulado y usa cadenas de Markov con algoritmo Monte Carlo (MCMC) para ajustar el modelo, lo que tiene el beneficio adicional de que las cantidades de interés a posterior (tal como, la probabilidad de que la unidad i – *esima* tenga un mayor valor de constructo latente con relación a la unidad j – *esima*) sean fáciles de calcular.

Modelo de análisis factorial para datos mixtos

Como en el análisis factorial estándar y en la teoría de ítem de respuesta, el objetivo del modelo es de captar los patrones de asociación entre las variables observadas a través de un modelo relativamente parsimonioso. Este tipo de modelo puede interpretarse también como variable latente en el caso de que los patrones de asociación observados surjan de una o unas variables inobservadas (es decir, latentes). Visto desde esta perspectiva, las variables de respuesta observadas son indicadores imperfectos de la(s) variable(s).

X Es una matriz $N \times j$ de respuestas observadas y cada variable observada puede ser ordinal o continua. Si la variable j – *esima* es ordinal tendrá $C_j > 1$ categorías. Se asume que los valores de los elementos de X son determinados por un matriz X^* de dimensión $N \times j$ de variables latentes y una colección de puntos de corte γ entonces:

$$x_{ij} = \begin{cases} X_{ij}^*, & \text{si la Variable } j \text{ es continua} \\ c, & \text{si } X_{ij}^* \in \gamma_{j(c-1)}; \gamma_{j(c)} \text{ y la Variable } j \text{ es ordinal} \end{cases} \quad (17)$$

Donde:

$j = 1, \dots, J$: Son los índices de las variables de respuesta

$i = 1, \dots, N$: Son los índices de las observaciones

c : Puede tomar valores de $\{1, 2, \dots, C_j\}$

Para identificar el modelo, se hace supuesto estándar de que $\gamma_{j0} \equiv -\infty$; $\gamma_{j1} \equiv 0$ y $\gamma_{jcj} \equiv \infty$ para todo j . Los puntos de corte restantes son parámetros libres a estimar.

Los patrones de asociación entre las variables observadas en \mathbf{X} son modeladas a través del factor analítico del modelo para la latente \mathbf{X}^* :

$$x_i^* = \Lambda \phi_i + \varepsilon_i \quad i = 1, 2, \dots, N \quad (18)$$

Donde:

x_i^* : Es el vector específico J de respuestas latentes para la i -ésima observación.

Λ : Es una matriz $j \times k$ de factor de cargas

ϕ_i : Es un vector específico k de puntuaciones de los factores para la i -ésima observación.

ε_i : Es vector J de perturbaciones que se \sim iid $N(0, \psi)$, ψ es diagonal.

Como se usan MCMC para ajustar el modelo, este se cambia ligeramente para acomodar la distribución condicional completa no *estándar* de los elementos libres de ψ . También será conveniente en algunos puntos de abajo escribir el sistema de N ecuaciones que se suministró en la ecuación anterior ($X_i^* = \Lambda \phi_i + \varepsilon_i$), como:

$$\mathbf{X}^* = \Phi \Lambda^t + E \quad (19)$$

Donde:

Φ : Es una matriz $N \times k$ con i filas iguales a ϕ_i

E : Es una matriz $N \times j$ con i filas iguales para ε_i

El primero elementos de ϕ_i es un set igual a 1 para todo i , esto asegura que los elementos en la primera columna de Λ funcionen como un ítems negativo para los

parámetros de dificultad de las variables de respuesta ordinal. Los elementos de la primera columna de Λ que corresponden a las respuestas continuas representan la media de estas variables continuas. Los elementos de la primera columna de Λ que corresponden a las variables continuas que han sido normalizados para tener media cero y deberán limitarse a cero. La estandarización de las variables continuas no es necesaria en este contexto, pero ayuda en la interpretación de los elementos de Λ que corresponden a las variables continuas y que pueden ser interpretados como factor de cargas.

Especificación a priori del modelo completo

Para identificar el modelo, algunos elementos de Λ puede ser restringidos a constantes, otros se verán restringidos a tomar sólo valores positivos (negativos), esto eliminaría la llamada invariancia rotacional. En general, para identificar el k-1 factor del modelo se debe permutar las filas de Λ de manera que:

$$\Lambda = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & 0 & \dots & 0 \\ \lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & & 0 \\ \lambda_{K,1} & \dots & & & & \lambda_{K,K} \\ \vdots & & & & & \vdots \\ \lambda_{J,1} & \dots & & & & \lambda_{J,K} \end{bmatrix}.$$

Al menos un elemento en cada columna de Λ debe estar restringido a tomar sólo valores positivos o negativos y se asume una densidad normal a priori recortada por debajo (o por encima) de 0, el resto de elementos de Λ siguen distribuciones normales independientes. Para ambos elementos normales atípicos a priori de Λ_{jk} recortados y no recortados, se asume que el parámetro media antes del recorte es μ_{0jk} y el parámetro precisión (varianza inversa) es L_{0jk} . Todos los elementos de Λ son asumidos como a priori independientes. Como es habitual en la literatura de la teoría de ítem de respuesta, los elementos de la diagonal de $\Psi = (\text{diag}\{\Psi_1^2, \Psi_2^2, \dots, \Psi_p^2\})$ que corresponden a las variables de respuesta ordinal están obligadas a ser igual a

1 por motivos de identificación y los elementos Ψ que corresponden a variables de respuesta continuas son dadas a priori como una distribución gamma inversa independiente, más específicamente si la variable j es continua $\Psi_{jj} \sim \text{IG}(a_{0j}/2, ab_{0j}/2)$.

Modelo de ajuste e inferencia

Con el software R para el análisis de datos y gráficos, en el paquete MCMCpack ese puede encontrar las funciones para el diagnóstico y resúmenes estadísticos en MCMC, el modelo que sigue el paquete se escribe en términos de la latente \mathbf{X}^* , es útil no sólo para construir una intuición, sino también para el ajustar del modelo. Basándose en las ideas del autor, se elige \mathbf{X}^* para tratar como datos latentes y se trabajar con la densidad posteriori.

$$\begin{aligned}
 p(\mathbf{X}^*, \gamma, \Lambda, \phi, \Psi | \mathbf{X}) &\propto p(\mathbf{X} | \mathbf{X}^*, \gamma) p(\mathbf{X}^* | \Lambda, \phi, \Psi) p(\gamma) p(\Lambda) p(\Phi) p(\Psi) \quad (21) \\
 &\propto \left\{ \prod_{i=1}^N \prod_{j=1}^J \left[\mathbb{I}(\mathbf{X}_{ij} = \mathbf{X}_{ij}^*) \right] (\mathbf{X}_j \text{ es continua}) \right. \\
 &\quad \left. + \sum_{c=1}^{c_j} \left[\mathbb{I}(\mathbf{X}_{ij} = c) \right] (\mathbf{X}_{ij}^* \in (\gamma_{j(c-1)}, \gamma_{j(c)})) \right] (\mathbf{X}_j \text{ es ordinal}) \left. \right\} \\
 &\quad \times p_N(\mathbf{X}_i^* | \Lambda \phi_i, \Psi) \left. \right\} p(\Lambda) p(\Phi) p(\Psi)
 \end{aligned}$$

Donde:

$\mathbb{I}(a)$: es la función del indicador, es igual a 1 si (a) es verdadero ó 0 en caso contrario.

$p_N(\mathbf{z} | \mu, \Sigma)$: es una densidad normal multivariada con media μ y matriz de varianzas-covarianzas Σ , evaluada en \mathbf{z} .

$p(\Lambda)p(\Phi)p(\Psi)$: son las densidades a priori de Λ, Φ, Ψ respectivamente.

La a prior para los puntos γ es una constante para todos los valores de γ .

La ventaja de trabajar con la parte a posteriori aumentada en la ecuación anterior, es que es relativamente fácil obtener distribuciones condicionales completas para la

mayoría de los parámetros del modelo a partir de esta densidad a posterior. Esto admite una muy simple cadena de Markov con algoritmo Monte Carlo (MCMC) que se utilizará para el ajuste del modelo, además, el algoritmo utilizado aquí funciona mediante un muestreador de las distribuciones condicionales completas de X^* , Λ , ϕ y Ψ y γ llamado *Metropolis-Hastings*.

El modelo presentado anteriormente proporciona un medio de principios para seguir adelante con el modelo basado en la medición de este tipo de situaciones donde hay conceptos latentes que tienen tanto variables continuas como ordinales y es fácil de usar e interpretar. También es posible extender el modelo presentado anteriormente en varias direcciones:

1) permite medir las correlaciones del error de forma bastante simple. Se requiere un Ψ a priori diferente y la simulación de Ψ determina otros parámetros del modelo que también cambiarán, pero el resto del algoritmo MCMC se mantiene intacto.

2) es posible permitir una gama más amplia de distribuciones con diferentes tipos de variables de respuesta. Por ejemplo, distribución Poisson con variables censuradas y/o variables amputadas, distribución exponencial y/o distribuciones gamma con variables continuas. Todas son todas posibles de acomodar en este marco general de modelización.

3) es posible extender el modelo para permitir la dependencia temporal, espacial o espacio-temporal. Esto es especialmente importante para aplicaciones de comparación y relación.

Diagnóstico convergencia de MCMC

El diagnóstico de convergencia en una cadena MCMC se usa para saber si se ha mezclado bien, para ello se utiliza un diagnóstico estándar donde se evalúa si la muestra de la cadena proviene aproximadamente de una distribución estacionaria. Las pruebas de convergencia más comunes son las propuestas por Gelman & Rubin (1992), Raftery & Lewis (1992) y el de Heidelberger & Welch (1983) que son en la

actualidad, los más populares, porque se apoyan en los programas informáticos para su cálculo:

El método de Gelman y Rubín (1992) se desarrolla en dos pasos, el primero se lleva a cabo antes del muestreo, para obtener una estimación más dispersa de la distribución de destino y generar de ella, los puntos de partida para el número deseado de cadenas y el segundo paso, se lleva a cabo para cada cantidad escalar de interés después de ejecutar las cadenas del muestreo de Gibbs para el número deseado de iteraciones y trata de utilizar las últimas dos iteraciones para volver a calcular la distribución de destino de la cantidad escalar como un conservador de la distribución t de Student .

El método de Raftery y Lewis, está propuesto para detectar la convergencia de la distribución estacionaria y para proporcionar una forma de delimitar las estimaciones varianza de los cuantiles de las funciones de los parámetros. Primero se ejecuta una sola cadena del muestreo Gibbs para el número mínimo de iteraciones que se necesitan para obtener la precisión deseada de la estimación si las muestras son independientes.

El método de Heidelberger y Welch, se emplea para la detección de no estacionariedad, en la salida de la simulación con un enfoque de análisis espectral para la estimación de la varianza de la media de la muestra, a través de un procedimiento general para la generación de un intervalo de confianza previamente especificado para la media cuando la simulación no ha empezado en esta distribución estacionaria y procedimiento debe aplicarse a una sola cadena. Aunque su enfoque es anterior al muestro de Gibbs, es aplicable a la salida del muestreo Gibbs y otros algoritmos MCMC.

1.7 ANÁLISIS DE VARIANZA

El análisis de varianza es una colección de modelos estadísticos y sus procedimientos asociados que permite determinar el efecto de distintos niveles de

alguna variable independiente X (variable explicativa) sobre alguna variable dependiente Y (variable explicada), en donde la varianza está particionada en ciertos componentes debidos a las diferentes variables explicativas. El análisis de varianza parte de unos supuestos que deben cumplirse como la independencia de las observaciones, la distribución normal de los residuales y la homocedasticidad u homogeneidad de las varianzas que se deben verificar mediante pruebas formales o gráficos antes de realizar el análisis de varianza (Terrádez, 2003).

Cuando se cumplen los supuestos y se realiza el análisis de varianza se pone a prueba la hipótesis referente a las medias de población más que a las varianzas de población a través del estadístico F de Fisher-Snedecor, un cociente que relaciona el efecto debido a la pertenencia de los grupos (variabilidad entre grupos) respecto a la dispersión debida al azar o error aleatorio (variabilidad dentro de los grupos). Si el valor del estadístico F calculado sobrepasa un valor tabular de F , indicará que el efecto observado es demasiado grande para poder ser explicado por el azar (error aleatorio) y por tanto no todos los grupos tendrán la misma media (Molinero, 2003).

Pruebas Post- Anova

Una vez se ha determinado la existencia de las diferencias entre las medias, las pruebas POST- ANOVA permiten determinar cuáles son las medias que difieren. Para ello, existen contrastes **no planeados** (cuando no hay idea previa de en qué grupos eran de esperar las mayores diferencias) y los contrastes **planeados** (Cuando se tienen sospechas de donde se deberían producirse las diferencias) (Baron & Tellez, 2011).

Para esta investigación se hará uso de una prueba que se clasifica en los contrastes no planeados llamada diferencia honestamente significativa de **TUKEY** o prueba **HSD-TUKEY**. Según Kuehl (2001), el método de Tukey proporciona la mejor protección contra errores de decisión, junto con inferencias fuertes sobre la magnitud y dirección de las diferencias con los intervalos de confianza simultáneos - ICS de $100(1 - \alpha)\%$. Esta prueba además, suele utilizarse cuando todos los tamaños de muestra son iguales, cuando se quiere comparar una gran cantidad de grupos y

cada uno de los grupo con todos los demás.

El conjunto de los métodos estadísticos expuestos en este apartado, serán los que nos permitirán tratar la información adquirida sobre las características agronómicas y organolépticas del café del Valle de Cauca, así como de las propiedades fisicoquímicas de los espressos obtenidos con diferentes perfiles de temperatura y tipo de café y finalmente, nos serán útiles para poder relacionar los descriptores organolépticos con las propiedades fisicoquímicas, así como para identificar los elementos discriminatorios para caracterizar un espresso.

1.8. PLANTEAMIENTO GENERAL

En el departamento del Valle del Cauca se está trabajando en mejorar cada una de las etapas de la cadena productiva del café con el objetivo de establecer cualidades o atributos comunes y no tan comunes que afectan la calidad global y que son determinantes para competir en el mercado (Muñoz, 2007; FNC, 2006 ; Ruiz et al. 2009). Sin embargo, en la actualidad no se dispone de una herramienta que permita seleccionar e identificar aquellas zonas o áreas cafeteras del departamento Valle del Cauca que se caracterizan por tener un café con buena calidad. Se requiere entonces, relacionar las características organolépticas del café mediante las pruebas de taza con las características del cultivo, recolección y beneficio, con el fin de generar un índice de calidad.

Si se lograra establecer un índice de calidad a partir de la relación entre las características organolépticas y las características del cultivo, recolección y beneficio, el departamento podría ser pionero en poder zonificar la calidad.

En cuanto al café espresso se sabe que la temperatura influye sobre las propiedades físicas y químicas de un tipo de café determinado ya que los diferentes componentes del café se solubilizan a distinta temperatura. La posibilidad de utilizar una máquina de café que trabaje a distintos perfiles de temperatura permitiría estudiar el efecto de la temperatura sobre las propiedades fisicoquímicas de un

determinado tipo de café. Así, se espera que la utilización de distintos perfiles de temperatura de extracción y un determinado tipo de café da lugar a espressos con distintas propiedades fisicoquímicas y por consiguiente a distintas propiedades organolépticas.

Además, la relación entre el análisis sensorial y las propiedades fisicoquímicas han de permitir identificar los elementos discriminatorios para caracterizar un café espresso.

2. OBJETIVOS

Las hipótesis planteadas conjuntamente con la revisión bibliográfica sobre los temas abordados en esta tesis, los métodos estadísticos existentes, así como los métodos de análisis físicos y químicos permitieron planificar cada uno de los estudios en la tesis y plantearse tres objetivos principales de los cuales se desprende la necesidad de conseguir algunos objetivos específicos.

1. Estimar el índice de calidad del café del Valle del Cauca en Colombia a partir de las características organolépticas y agronómicas empleando Análisis Factorial Bayesiano.
 - Obtener un archivo de datos completo y consistente a través de la evaluación y selección de una de tres técnicas de imputación de datos (Listwise, hotdeck e imputación múltiple).
 - Estimar el índice de calidad del café a través del Análisis Factorial bayesiano mediante MCMC (Cadenas de Markov Monte Carlo).
 - Calcular las cargas factoriales y el porcentaje de varianza explicada, para cuantificar la contribución al factor común (índice de calidad) de las características organolépticas y agronómicas.
 - Presentar la zonificación geográfica del índice de calidad y las áreas con mayor probabilidad de este índice.
2. Determinar la influencia en las propiedades físico-químicas de los espressos

obtenidos con tres cafés tostados utilizando distintos perfiles o curvas de temperatura de extracción mediante el análisis de varianza.

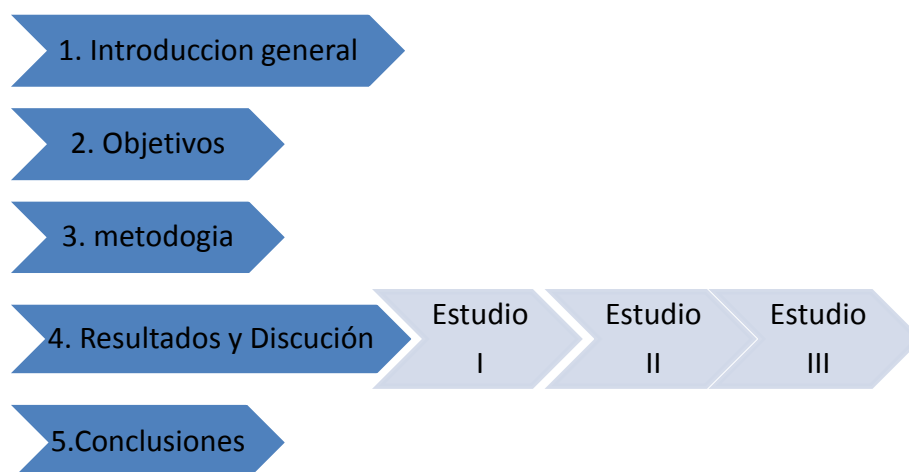
- Poner a punto las técnicas analíticas para obtener las propiedades fisicoquímicas del café espresso.
 - Determinar las propiedades físicas y químicas de los espressos obtenidos con la máquina de café XCELSUS Classe 9 a partir de tres cafés tostados (Robusta Natural, Arábica Natural y Arábica Lavado) modificando el tamaño de molido según el punto de tueste del café y las curvas o perfiles de temperatura 90-90 °C, 88– 93 °C y 93–88 °C.
 - Señalar las diferencias estadísticas de las propiedades fisicoquímicas en los tres tipos de café en función del perfil de temperatura.
 - Relacionar las propiedades fisicoquímicas de los espressos con los descriptores organolépticos.
3. Identificar los elementos discriminatorios para caracterizar un café espresso empleando un Análisis de Componentes Principales.
- Investigar cuáles son los principales descriptores organolépticos que contribuyen a la calidad del café espresso preparados con Robusta Natural, Arábica Natural y Arábica Lavado en función de los perfiles o curvas de temperatura de extracción (90-90 °C, 88– 93 °C y 93–88 °C).
 - Determinar las características fisicoquímicas asociadas a los descriptores organolépticos.
 - Contribuir a la interpretación de la calidad del café y su relación con los descriptores organolépticos.

3. METODOLOGÍA

Introducción

De acuerdo al objeto de estudio de esta investigación y la perspectiva estadística desde la cual se aborda, se propone dos enfoques metodológicos desde el cual se hace uso de la estadística multivariada y el diseño experimental para llegar al objetivo de poder hallar los índices de calidad del café a través de las características organolépticas y fisicoquímicas.

Una síntesis del procedimiento propuesto se presenta en la Figura 2, como se observa, esta investigación se estructura en cinco etapas que corresponden a una Introducción general, objetivos, la metodología, los resultados y discusión y finalmente las conclusiones, a continuación se presenta una breve descripción.



Estudio I: Índice de calidad del café del Departamento del Valle del Cauca en Colombia a partir de las características organolépticas y agronómicas. **Estudio II:** Calidad del café espresso a partir de propiedades fisicoquímicas empleando perfiles de temperatura de extracción. **Estudio III:** Identificación de elementos discriminatorios para caracterizar el café espresso.

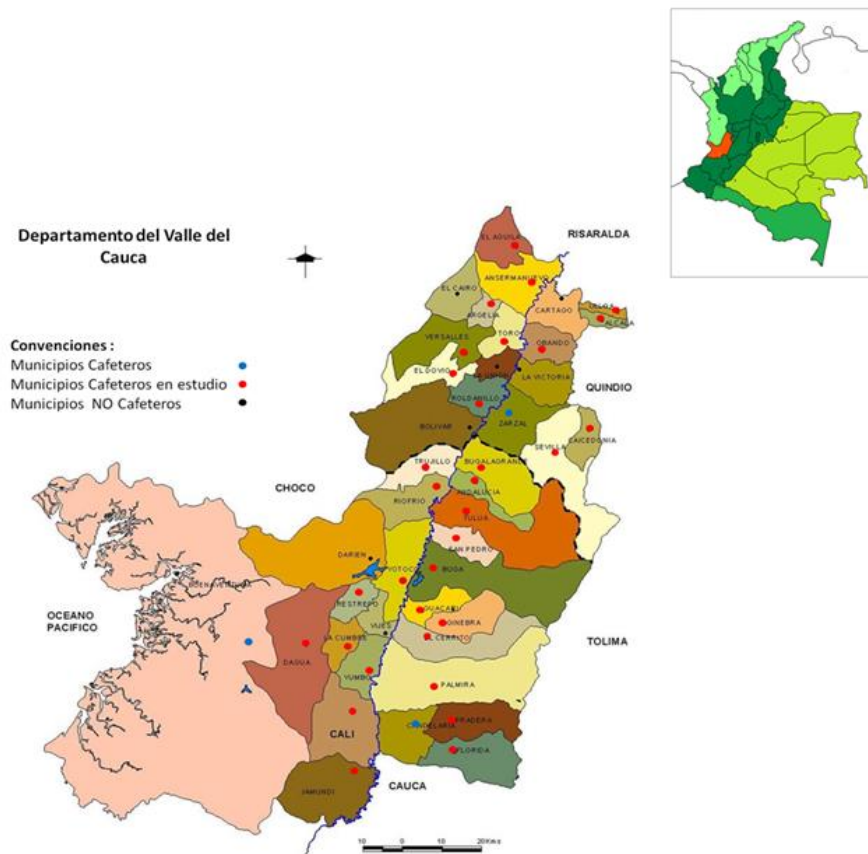
Figura 2. Estructura de la tesis.

En la primera parte, tal y como se expresa en la revisión de literatura que sustenta esta investigación, se hace uso de diferentes técnicas estadísticas cuyo análisis, busca contrastar los supuestos iniciales y los objetivos que dan origen a este proyecto. En la segunda parte, se hace una descripción metodológica detallada del trabajo en general y de cada uno de los estudios. En la tercera parte que corresponde a los resultados y discusión, se dividen especialmente en los tres estudios: 1) *Índice de calidad del café del Valle del Cauca en Colombia a partir de las características organolépticas y agronómicas empleando análisis factorial bayesiano*, 2) *propiedades fisicoquímicas de tres cafés tostados mediante el uso de la máquina de café espresso, modificando el tamaño del molido y la temperatura de extracción* y 3) *identificación de elementos discriminatorios para caracterizar el café espresso empleando un análisis de componentes principales*. Cada uno intenta responder a los objetivos propuestos. Finalmente, se concluye este trabajo con las conclusiones de cada uno de los estudios descritos previamente.

3.1. ÍNDICE DE CALIDAD DEL CAFÉ DEL VALLE DEL CAUCA EN COLOMBIA A PARTIR DE LAS CARACTERÍSTICAS ORGANOLÉPTICAS Y AGRONÓMICAS

Ubicación geográfica del estudio

La base de datos fue proporcionada por el comité de cafeteros del Valle del Cauca y fue elaborada con 900 muestras enviadas por los caficultores de 32 de 39 municipios cafeteros del departamento del Valle del Cauca, durante el año 2009-2010, las cuales llegan identificadas con los datos de la finca y el proceso hecho al café. Los municipios cafeteros se muestran en la Figura 3, con un punto de color rojo.



Fuente: IGAC-CVC-FEDECAFE

Figura 3. Mapa del Valle del Cauca con sus municipios cafeteros incluidos en el estudio.

Obtención de la Información

La Federación Nacional de Cafeteros de Colombia - FNC diseño 3 formatos: el **formato de etiqueta** para identificar la muestra, el **formato de análisis físico del café** para registrar los resultados físicos y él **formato de prueba de taza** basado en la metodología de la Asociación de Cafés Especiales de América (SCAA) que se emplea para el análisis sensorial de los cafés especiales. A partir de estos tres formatos se creó la base de datos con 900 pruebas de taza del departamento del Valle del Cauca con el cual se desarrolló este estudio.

Características de las variables

Las variables definidas en el estudio se dividen en tres grandes grupos, las variables de análisis sensorial, las variables de análisis físico de la calidad del café y las variables espaciales y agronómicas, cabe resaltar que la naturaleza de estas variables es de carácter discreto, continuo y nominal categórica según sea el caso. Cada uno de estos grupos se describe a continuación.

Análisis Sensorial

En la tabla 2, se muestran las variables organolépticas del café, que se obtuvieron del formato de prueba de taza. La escala es de 6 a 10 para cada atributo y hasta 100 en la puntuación final.

Tabla 2. Variables del análisis sensorial de las pruebas de taza que son resumidas en la puntuación final.

	Descripción y Calificación
Fragancia	Olor del café en seco. 0 a 10 (10 es el punto máximo)
Aroma	Olor del café una vez se adiciona el agua. 0 a 10 (10 es el punto máximo)
Sabor	Impresión combinada de todas las sensaciones gustativas. 0 a 10 (10 es el punto máximo)
Sabor Residual	Sabor que queda en la boca después de probado el café. . 0 a 10 (10 es el punto máximo)
Acidez	Contribuye a la vivacidad del café, dulzor y al carácter de fruta fresca. . 0 a 10 (10 es el punto máximo)
Cuerpo	Sensación táctil del líquido en la boca. 0 a 10 (10 es el punto máximo)
Balance	Equilibrio entre acidez, cuerpo, sabor y dulzor. 0 a 10 (10 es el punto máximo)
Taza Limpia	Es la transparencia de la bebida, sin defectos. 2 puntos por taza.
Dulzor	Plenitud agradable del sabor en la boca. 2 puntos por taza.
puntuación catador	Impresión global del café. 0 a 10 (10 es el punto máximo)
Defectos	Sabores negativos. Se restan 2 puntos por taza si es ligero y 4 si es rechazo.
puntuación Final	Es la suma de los atributos menos los defectos. Calificación Máxima 100 puntos

Procedimiento para la cata

El análisis sensorial de todas las muestras de café de todos los municipios del Valle del Cauca, se realizó mediante un grupo de tres catadores que han sido seleccionados y capacitados mediante diferentes pruebas sensoriales y concursos. El método está descrito por Martínez (2013) y se denomina cata de café a la brasileña. El objetivo de este procedimiento es analizar las características organolépticas de una forma simple y sin interferencias. Para ello, se comenzó por tostar la muestra en el punto de tueste claro a claro-medio, se dejó reposar por 8 horas como mínimo para que se enfriara a temperatura ambiente y se guardó en bolsas herméticas alimentarias. Después de este tiempo se pesó 8.5 gr de café en grano por muestra y se molió inmediatamente antes de la cata. Cada muestra se depositó en la taza y a continuación se agregó aproximadamente 150 mL de agua caliente a 93 °C en el momento de llenar la taza. Posteriormente se dejó reposar de 3 a 5 minutos antes de ser evaluado y la información se almacena en el formato. Se realizaron 5 pruebas de taza por cada muestra de café.

Para los resultados de este análisis se excluyeron los datos de uniformidad, taza limpia y dulzor, ya que estos tres datos, muestran si es o no limpia la taza y si hay o no igualdad entre las pruebas; además se aclara que con respecto al dulzor solo se mencionó si está presente o no, ya que no está calificado ni por intensidad, ni calidad.

Análisis físico del café

Las variables se obtuvieron del formato de registro de análisis de la calidad física del café una vez realizada las pruebas al café pergamino seco – CPS y a la almendra, bajo los estándares de la Federación Nacional de Cafeteros de Colombia para la comercialización del CPS y los procedimientos del laboratorio de Calidad y Manejo Ambiental de Cenicafé. A continuación se muestran las variables y una breve descripción.

Tabla 3. Variables de análisis físico de calidad de café.

Variable	Descripción y Calificación
Humedad en Café Pergamino Seco - CPS	Cantidad de agua presente en el CPS, importante para la calidad y la comercialización. 10 % - 12%.
Aspecto del CPS	El pergamino debe presentar uniformidad en color y debe estar libre de olores extraños o de cualquier tipo de contaminación. Color normal. Olor fresco característico a olores distintos (como a reposo, moho, tierra, vinagre, productos derivados del petróleo, etc.)
Almendra sana	Almendras sin efectos físicos. Sin presencia de granos con pulpa e impurezas. 75% - 100 %
Merma en trilla	Denominación que se le da a la pérdida del peso del grado de café mediante el proceso de descascarado para obtener el café excelso o almendra. 0% - 20 %.
Almendra defectuosa	1. Cardenillo, 2. Cristalizado, 3. Decolorado (blanqueado, ámbar o mantequilla), 4. Manchado, 5. Mordido, cortado, 6. Picado por insectos (afectado por la broca), 7 Partido, 8. Malformado o deformado, 9. Inmaduro, 10. Aplastado, 11. Flotador o balsudo, 12 Flojo, 12. Negro balsudo, 14. Vano, 15. Astillado y partido, 16. Reposo, 17. Sucio, 18. Sabor fenólico, 19. Materias extrañas. 0% - 5%.
Color almendra	Característico, cutícula ligeramente rojiza, cutícula rojiza, decolorada, decolorada vetada, dispareja y sobresecada.
Olor almendra	Normal, agrio, cigarrillo, fermento, moho, tierra vinagre

Variables espaciales y agronómicas

Las variables espaciales y agronómicas del cultivo del café en campo y poscosecha provienen del formato de etiqueta que se diligencia en la finca y con el técnico quien verifica la información a través del sistema de información cafetera-SICA. La tabla 4. Presenta las variables, su descripción y su unidad de medida.

Tabla 4. Variables Espaciales y Agronómicas del estudio.

Variable	Descripción	Unidad de Medida
Municipio	Municipios objeto de estudio	Cali, Alcalá, Andalucía, Ansermanuevo, Argelia, Bolívar, Buga, Bugalagrande, Caicedonia, Dagua, El Águila, El Cerrito, El Dovio ,Florida, Ginebra, Guacarí, Jamundí, La Cumbre, Obando, Palmira, Pradera, Restrepo, Riofrío, Roldanillo, San Pedro, Sevilla, Toro, Trujillo, Tulúa, Ulloa, Versalles, Yótoco y Yumbo
Variedad	Variedades de café analizadas en las pruebas de taza	Castillo, Colombia, Catimor, Caturra, Típica, Tabí, Costa Rica
Altura (msnm)	Altura de siembra del cultivo sobre el nivel del mar	1000 msnm - 2000 msnm
Fertilización	Fertilización empleada en el cafetal	Convencional, Natural, Orgánica, ninguna
Sombrío	Tipo de exposición del café al sol	Sombrío, al sol
Beneficio	Tipo de beneficio empleado	Desmucilaginado, fermentación natural, Desmucilaginado y fermentación natural.
secado	Secado del café	Al sol, mecánico, Al sol y mecánico

Descripción del método estadístico

Imputación de Datos

Ante la presencia de la información faltante en algunas variables en la base de datos en estudio, se recurrieron a técnicas de imputación de datos que básicamente radican en completar la base de datos bajo cierto supuestos, ya sea eliminando los casos con datos faltantes o reemplazando dichos datos por otros, tan similares a ellos como sea posible. Para ello se realizaron tres técnica de imputación de datos, Listwise o case deletion - CD Imputación Múltiple, - MI y Hot Deck - HD que permitieron crear bases de datos completas, donde se analizó mediante métodos estadísticos tradicionales (valores de la media (Me), mediana (Med), desviación estándar (SD), coeficiente de asimetría (As), curtosis (K) y el contraste de normalidad de Kolmogorov-Smirnov-Lilliefors (KSL)) con el objetivo de evaluar las técnicas de imputación a través de la simulación y poder seleccionar la más adecuada. La

técnica de MI se escogió como la técnica de referencia, debido a que según Rubín (1987) y otros autores toma en cuenta una cantidad considerable de valores perdidos y no informados (NA/NI) en algunas de las variable y además aquellas que presentan una elevada variabilidad de los valores imputados que según Schafer (1997), es conveniente tratar con esta técnica.

Análisis Factorial

La base de datos seleccionada y los datos iniciales fueron analizados mediante el análisis factorial bayesiano propuesto por Quinn (2004), que basa su teoría en la utilización de las cadenas de Markov con algoritmo Monte Carlo (MCMC) para estimar el índice de calidad objetivo central de este estudio. Posteriormente, se mostraran las zonas geográficas que presentan una mayor probabilidad de este índice de calidad. Se calcularon las cargas factoriales y el porcentaje de varianza explicada necesarias para cuantificar la contribución del factor común (índice de calidad) a las características organolépticas y agronómicas. Además se verificó la convergencia de las MCMC mediante la prueba de Heidelberger y Welch (1983).

Todo el procesamiento de la información se hizo en el software R, a través de diversos paquetes estadísticos entre ellos los más destacados:

STATS: Listwise o case deletion CD

VIM: Hot Deck -HD

MICE: Imputación Múltiple – MI

MCMCpack: Análisis factorial bayesiano

3.2. CALIDAD DEL CAFÉ ESPRESSO A PARTIR DE LAS PROPIEDADES FÍSICOQUÍMICAS EMPLEANDO PERFILES DE TEMPERATURA DE EXTRACCIÓN

Los datos sobre las propiedades fisicoquímicas del café espresso fueron obtenidos en el laboratorio de Análisis Alimentario del departamento de Ingeniería Química,

Agraria y Tecnología Agroalimentaria de la Universitat de Girona, en Catalunya que es el lugar donde se desarrolló la parte experimental de esta tesis. La máquina de moler café, así como la máquina para hacer las extracciones o muestras de café espresso fueron proporcionadas por la empresa Rancilio Group.

Establecimiento de las condiciones de laboratorio

Para obtener análisis uniformes en términos de calidad de la muestra, se establecieron previamente algunos aspectos importantes para el desarrollo del estudio que se mencionan a continuación:

Calidad del Agua

El agua mineral empleada durante todo el estudio presentó la composición química que se muestra en la Tabla 5. Como se observa, el agua empleada presenta una baja mineralización ya que presenta una cantidad razonable de bicarbonatos y calcio.

Tabla 5. Calidad de Agua.

Composición	mg/L	Composición	mg/L
Bicarbonatos	113	Potasio	4,9
Calcio	27,7	Residuo seco	139
Cloruros	5,7	Sodio	11,9
Flúor	0,9	Sulfatos	11,2
Magnesio	4,5		

Fuente: Viladrau (2013)

Tipos de Café

Se seleccionaron tres diferentes tipos de café según su especie y el tipo de beneficio (seco o húmedo). El punto de tueste de los cafés se definió como un 7/10 Dark medium y se les asignó la siguiente nomenclatura que se describe en la Tabla 6.

Tabla 6. Clasificación del tipo de café

Café	Etiqueta	Especie	Beneficio del café	Origen
Tipo A	Robusta Natural	Robusta	seco	Vietnam
Tipo B	Arábica Natural	Arábica	seco	Brasil
Tipo C	Arábica Lavado	Arábica	húmedo	Perú

Procedimiento de extracción y calibración de la máquina

La máquina para obtener los extractos de café, es la Xcelsius clase 9 de la empresa Rancilio la cual se muestra en la Figura 4. Esta máquina se compone de tres grupos de extracción (1, 2 y 3); de cada grupo, se obtiene 2 tazas de café espresso (CE). Todos los grupos fueron calibrados para tener controlados los parámetros de presión (9 bares) y los perfiles o curvas de temperatura (90-90 °C, 88-93 °C y 93-88 °C) deseados. Una vez fijada la curva de temperatura deseada, se ajustó el punto de molido de cada uno de los tres tipos de café para obtener un café espresso - CE en el tiempo y volumen deseado, según las normas de calidad para obtener un auténtico café espresso.



Figura 4. Máquina Xcelsius clase 9 de la empresa Rancilio.

Preparación de la muestra

Cada extracción de café se efectuó pesando previamente 15 ± 0.5 g de café molido por grupo de la extracción donde se obtuvieron dos CE. El tiempo fue controlado

25 ± 5 en segundos y el volumen en 25 ± 2.5 mL de CE por taza según el Instituto Nazionale Espresso Italiano (2013).

Propiedades físicas y químicas del café espresso – (CE)

Una vez obtenidos los extractos se llevaron a cabo las determinaciones de las propiedades físicas y químicas que se mencionan a continuación:

El pH y la densidad. EL pH en las muestras de las extracciones de CE se midió en un pHmetro (Crison micro pH 2000), la densidad fue medida con un picnómetro Gay-Lussac de 25 mL.

El Volumen de la crema y su persistencia. El volumen de la crema y del líquido se midió inmediatamente después de obtener el CE en un vaso de precipitados de 50 mL de marca Brand de (Ø=42 mm y altura de 60mm) previamente calentado a 31 °C. La altura de la crema y de café se midió con un vaso de escala milimetrada. La persistencia que es el tiempo transcurrido inmediatamente después de ser preparado el CE hasta ver la primera ruptura de la crema y observar la fase líquida en el extracto a temperatura ambiente se determinó a través de una medida cronométrica (en minutos y segundos). Estas dos propiedades se midieron en caliente.

Volumen del extracto y viscosidad: Una vez enfriado el CE se midió su volumen mediante una probeta graduada de 50 mL. La viscosidad se midió a través del viscosímetro de Ostwald (PROTON 100) teniendo en cuenta la temperatura del día de la medición.

Sólidos totales y sólidos filtrados. Los sólidos totales se determinaron por el secado de 10 mL de CE en una estufa de laboratorio a 105 °C por 24 horas o hasta conseguir un peso constante. La medida fue definida como g/L de café con respecto a la dosis de café tostado molido (7,5 g). Los sólidos filtrados se determinaron por el secado en la estufa de 10 mL del CE después de ser filtrado con un papel filtro de microfibras de vidrio 1.2 µm FILTER-LAB que se llevó a la estufa a 105 °C por 24 horas para igualmente conseguir un peso constante.

Turbidez. Se midió por turbidimetría (mide la disminución de la transmitancia del haz de luz al atravesar la muestra) en un equipo de turbidez (HANNA LP2000). Para realizar la medida en el rango de medición del instrumento se diluyeron 0,4 mL del CE en 100 mL de agua.

Lípidos totales. La determinación total de los lípidos se realizó por extracción líquido-líquido utilizando hexano. El procedimiento se hizo midiendo 50 mL del CE del cual se extrajeron los lípidos mediante la adición de 20 mL de hexano en un embudo de decantación. El proceso de extracción con hexano se realizó tres veces. La fracción orgánica se lavó con 60 mL de agua destilada cinco veces, y se utilizó sulfato de sodio para eliminar los restos de agua de la fracción orgánica. Se filtró la solución para separar la fracción sólida. Los lípidos se cuantificaron mediante el peso después de la evaporación del disolvente (Ratnayake et al.1993).

Polifenoles totales. Se siguió el método de Folin-Ciocalteu (Singleton & Rossi, 1965), basado en la oxidación de los grupos hidroxilos de los compuestos fenólicos en medio básico mediante el reactivo de Folin-Ciocalteu. La determinación de polifenoles totales se realizó por detección espectrofotométrica en un equipo Gilson 223 Sample Changer (dilutoautomàtic "Ciocalteu reagent DC"). Se colocaron 2 mL de CE en tubos de ensayo que después se colocaron en el equipo, este lo diluye la muestra 12.5 veces con 600 μ L Na₂COO₃ y lo enrasa con agua Milli-Q hasta llegar a 6 mL. Seguidamente se dejó reposar 2 horas y se midió la absorbancia de las muestras a 750 nm. Se realizó el mismo procedimiento con distintos patrones de ácido gálico de 200, 400, 600, 800, y 1000 mg/L para obtener la recta de calibración. Los resultados de absorbancia se cuantificaron con la recta de calibración y el contenido en polifenoles totales se expresó como equivalentes en ácido gálico.

Color. Las lecturas de intensidad de color se hicieron espectrofotométricamente midiendo la absorbancia de la solución a través de un espectrofotómetro A-Visible Espectrofotómetro UV-2450, Shimadzu®, Kyoto, Japón. Para ello, se prepararon diluciones de los extractos de café llevando 0.4 mL de CE a 100 mL con agua Milli-Q en un matraz aforado. Las medidas se realizaron a dos longitudes de onda distintas,

a 465 y 665 nm. Se expresó la intensidad de color como la relación entre la absorbancia a 465 y a 665.

Cafeína. El análisis se llevó a cabo utilizando una unidad de HPLC analítica (KONIK HPLC 550 A) equipada con una columna Nucleosil 100 C18 (5 μm tamaño de partícula, 100x4.6 mm) y con un loop de inyección de 20 μL . El caudal de fase móvil por la columna fue de 0.8 mL/min. La fase móvil fue metanol / agua (40:60) y la detección se realizó a 210 nm. Se analizaron patrones de cafeína de 5, 15, 10, 20 y 50 mg/L para obtener la curva de calibración. Previamente al análisis, se diluyó 0,4 mL del CE con agua Milli-Q en un matraz aforado de 100 mL y antes de la inyección las muestras se filtraron con un filtro de nylon de 0.2 μm y posteriormente se procedió a la determinación de la cafeína.

Ácido clorogénico. La determinación de los ácidos clorogénicos se realizó mediante el HPLC (Agilent 1200) equipado con una columna Eclipse XDB C-18 (5 μm tamaño de partícula, 250 x 4.6 mm) con una velocidad de flujo en la columna de 0.8 mL/min, un detector de longitud de onda múltiple, y un loop de inyección de 20 μL . Se realizó la cromatografía con un gradiente de fase móvil formada por dos soluciones A y B. La solución A es agua con 1% de ácido fórmico y la fase B acetonitrilo y 0.1% de ácido fórmico. La detección se hizo a una longitud de onda de 325 nm.

Con el fin de eliminar las proteínas y las moléculas de gran tamaño se hizo un pre-tratamiento a la muestra antes de su análisis con HPLC. Este tratamiento consistió en tomar 8 mL de CE, añadir 0.2 mL de la solución reactivo Carriez I (250 mM de ferrocianato de potasio tetrahidratado), 0.2 mL de la solución de reactivo Carriez II (1M acetato de zinc dihidratado que contiene 3% de ácido acético glacial (V/V), y 1.6 mL de metanol. Después de agitar durante 5 segundos, se dejó reposar la mezcla 10 min y se centrifugó durante 10 minutos a 5000 rpm en una centrífuga (Hettich Universal 320). Se tomó una muestra 0.2 μm del sobrenadante previamente filtrado y se analizó por cromatografía. Para poder cuantificar el ácido clorogénico se realizó una recta de calibración con patrones de 0, 50, 100, 150, 200, 250 y 500 mg/L, preparados a partir de la disolución de 0.01 g de ácido clorogénico 5-cafeoilquinico

(5-CQA) en 100 mL de agua Milli-Q. Los patrones se analizaron bajo las mismas condiciones cromatográficas utilizadas para el análisis de las muestras (Trugo & Macrae, 1984).

Granulometría o tamaño de partícula

Para determinar la distribución de partículas, inicialmente se establecieron las curvas de temperatura de extracción en la cafetera y el punto de molido de cada uno de los cafés como se explicó anteriormente y se pesó 10 g de café molido según esta configuración. En la medición se emplearon tres diferentes tamaños de tamiz con diámetros de 500 μm 250 μm y 150 μm accionados por una máquina de tamizar de marca CISA BA 200N fabricada en España (ajustada entre 28-30 movimientos por minuto) en donde la cantidad de café en masa que queda retenido en cada tamiz se relaciona con el total de la masa de la muestra analizada y se expresa como porcentaje. Posteriormente la molienda se clasificó como gruesa (>500 μm), media (500-250 μm) y fina (< 150 μm) (Guevara & Castaño, 2005; IICA, 2010).

Descripción del diseño y análisis del método estadístico

Análisis de varianza

El diseño estadístico implementado fue un diseño de dos factores, el tipo de café (A = Robusta Natural, B= Arábica Natural y C= Arábica Lavado) y el perfil o curva de temperatura de extracción (ascendente= 88-93 °C, plana= 90-90 °C y descendente= 93-88 °C) donde se rechazó la hipótesis de que todas las medias son iguales cuando se presentó una significancia baja (< 0.05).

Cada una de las extracciones se realizó con 7 repeticiones por grupo y se descartó siempre la primera extracción para evitar interferencias. Los extractores de la máquina se limpiaron después de cada cambio de tratamiento. En total se realizaron 18 repeticiones por tratamiento.

Una vez obtenidos los extractos, se realizaron las determinaciones de laboratorio y

los resultados se analizaron preliminarmente mediante estadísticas descriptivas para observar el comportamiento de las variables.

Posteriormente se realizaron los análisis de varianza (ANOVA) para analizar la variación entre muestras y la variación al interior de las mismas y finalmente se efectuó la prueba post hoc.

Pruebas post hoc

La prueba post hoc utilizada para la diferencia de medias o de comparaciones múltiples en aquellas variables estadísticamente significativas fue el test HSD (Honestly-significant-difference) de Tukey con un nivel de significancia del 10%.

Se empleó el test de Tukey, ya es una de las pruebas más potentes cuando se contrasta un gran número de pares de medias porque conduce a límites de confianza de menor longitud encontrando diferencias significativas más óptimas, esto porque utiliza el estadístico del rango estudentizado para realizar todas las comparaciones por pares entre los grupos y establece la tasa de error por experimento como la tasa de error para el conjunto de todas las comparaciones por pares.

Todos los análisis estadísticos de los utilizados durante el desarrollo esta tesis se realizaron mediante el software estadístico R.

3.3. ELEMENTOS DISCRIMINATORIOS PARA CARACTERIZAR EL CAFÉ ESPRESSO

El estudio de la identificación de los elementos discriminatorios para caracterizar el café espresso, se llevó a cabo en el laboratorio de Análisis Alimentario del departamento de Ingeniería Química, Agraria y Tecnología Agroalimentaria de la Universitat de Girona, en Catalunya que es el lugar donde se desarrolló la parte experimental de esta tesis. La máquina de moler café, así como la máquina para

hacer las extracciones o muestras de café espresso (CE) fueron proporcionadas por la empresa Rancilio Group.

Además, en este laboratorio se realizaron las pruebas de taza para determinar los descriptores organolépticos de los CE, a través un panel de catadores provenientes de diversas empresas reconocidas en la importación y comercialización del café en España.

Obtención de los cafés espressos

Los extractos de café espresso para la determinación de las propiedades fisicoquímicas, así como para los descriptores organolépticos fueron preparados por un mismo barista con el objetivo de no alterar los resultados, del mismo modo, se controlaron los parámetros de la extracción (cantidad, calidad y temperatura del agua) y el tiempo de erogación siguiendo el procedimiento de extracción y calibración de la máquina, así como la preparación de la muestra que se describe en el apartado 3.2. Cabe anotar que se utilizaron siempre los mismos tres tipos de café y perfiles de temperaturas realizadas en el estudio anterior.

Variables del estudio

Las variables del estudio fueron agrupadas en dos conjuntos: las fisicoquímicas que provinieron de los análisis fisicoquímicos y las organolépticas que provinieron de la cata de los espressos. A continuación se hace una breve descripción de los análisis.

Análisis fisicoquímicos

Los análisis fisicoquímicos fueron realizados conforme al protocolo establecido anteriormente y detallado en el apartado de 3.2 en propiedades físicas y químicas del café espresso (CE). Se conservaron las mismas variables menos la densidad del café, que fue eliminada por no haber presentado diferencias estadísticas significativas.

Análisis sensorial

El análisis sensorial del café expreso se realiza para el café terminado, tal y como se va a consumir en casa o en la cafetería. Las muestras de CE fueron evaluadas por seis catadores entrenados de diferentes empresas reconocidas en la importación, tueste y distribuciones del café, mediante valoraciones visuales, olfativas, gustativas y retroolfativas.

Establecimiento de las condiciones para la cata

Como se mencionó anteriormente la cata del espresso se realizó en el laboratorio bajo condiciones adecuadas de espacio, luz, limpieza, temperatura entre 20-24 °C y un ambiente silencioso para facilitar la concentración de los catadores. Además se emplearon cubículos para evitar la interferencia entre panelistas. Las catas se realizaron entre las 10 y 12 de la mañana, ya que en este momento del día las papilas gustativas están mejor preparadas.

Procedimiento para la cata del espresso

La descripción sensorial se realizó de acuerdo al protocolo de cata de espresso descrito por Cuadras (2009) y el cual tiene el objetivo de evaluar el conjunto global del blend u origen y valorar su comportamiento en la taza. El procedimiento comenzó con la selección de las tazas de café expreso en porcelana, de color blanco, redondeadas, de fondo cóncavo, con una capacidad media de 50 mL y precalentadas a 31 °C.

A continuación se realizaron las extracciones de café con las condiciones previamente descritas y se procedió a efectuar la evaluación sensorial, dando aleatoriamente a cada panelista un CE recién elaborado y una hoja de registro con los diez descriptores organolépticos que se midieron en una escala de 0 a 10. En la Tabla 7. Se muestra el descriptor y sus características.

Tabla 7. Descriptores organolépticos

Descriptor	Características
Intensidad Color Crema	Referencia al color avellana con tonos atigrados.
Textura de la crema	Crema densa (burbujas muy pequeñas, casi invisibles), con ninguna burbuja de aire ni manchas claras y sin insinuar la tendencia a abrirse dejando entrever el líquido del fondo.
Intensidad Olfativa	Fuerza del olor (que puede ser agradable o no)
Intensidad Tostado	Resalta las notas gustativas. Si el tueste es ligero resalta notas gustativas cítricas frutales, florales y miel) si es medio resalta sabores a nuez, chocolate y caramelo) y si es oscuro resaltan los sabores chocolate amargo.
Cuerpo	Sensación de perdurabilidad, intensidad, redondez y espesor en la boca.
Acidez	Referencia a la frescura o suavidad y viveza. Entre más suave el café es más ácido.
Amargo	Se siente en toda la boca e incluye la lengua (Referencias: solución de quinina o cafeína).
Astringencia	Sensación compleja olfato-gustativa, generalmente debido a la presencia de compuestos ácidos. Acidez no franca.
Olores +	Importancia de la nota característica del olor de café.
Olores-	Aspectos negativos en el olor.

Descripción del método estadístico

Estadísticas descriptivas

Para el análisis de la información, inicialmente se analizaron estadísticas tales como media, desviación estándar y coeficiente de variación de todas las variables fisicoquímicas y de los descriptores organolépticos discriminado los espresso provenientes de extractos hechos con los tipos de café (A = Robusta Natural, B= Arábica Natural y C= Arábica Lavado) a diferentes perfiles o curvas de temperatura de extracción (ascendente=88-93 °C, plana=90-90 °C y descendente=93-88 °C).

Análisis de Componentes principales - ACP

Para el análisis de los descriptores organolépticos se empleó un método estadístico multivariado exploratorio denominado ACP, en el cual se parte de una matriz Y de dimensiones $n \times p$, donde p corresponde al número de variables observadas de los descriptores organolépticos en n muestras de café conformadas por el perfil de temperatura de extracción y el tipo de café.

Antes de aplicar el ACP se comprobó si era necesario hacerlo. Para ello, se realizó el test de Bartlett y se halló el índice de Kaiser-Meyer-Olkin o medida de adecuación muestral KMO para justificar la factorización de la matriz de coeficientes de correlación.

Finalmente, se realizó un ACP empleando los descriptores organolépticos por perfil de temperatura y el resultado obtenido de cada ACP se sintetizó en dos representaciones gráficas (Duarte et al. 1996; Pérez, 2004).

Representaciones graficas del ACP

El diagrama de variables o círculo de correlación, en el que se representan las variables en el plano factorial retenido que corresponden a los coeficientes de correlación entre las variables originales (descriptores organolépticos) y las componentes; y el diagrama de muestras, en el que se representan las muestras de café u observaciones en el plano factorial. (Smith, 2002; Anderson, 2003b).

En total, fueron 72 muestras de café espresso y 17 variables contempladas en el análisis del estudio, para los ACP se emplearon solo 8 variables que corresponden a los descriptores organolépticos. Todo el análisis estadístico de los datos fue realizado con el programa R.

4. RESULTADOS Y DISCUSIÓN

4.1 ÍNDICE DE CALIDAD DEL CAFÉ DEL VALLE DEL CAUCA EN COLOMBIA A PARTIR DE LAS CARACTERÍSTICAS ORGANOLÉPTICAS Y AGRONÓMICAS.

4.1.1 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES

En las Tablas 8 y 9, se muestran las estadísticas descriptivas para las variables del estudio tanto cuantitativas como cualitativas que formaron parte del estudio.

Tabla 8. Variables de los datos originales (sin imputación).

Variables	Etiquetas	N		Media	Mediana	Mínimo	Máximo
		Válidos	Faltantes				
Variedad	varied	692	299	10,06	12	1	15
Sombrío	sombr	649	342	9,30	5	1	27
Fertilización	fertiliz	753	238	2,93	1	1	3
ALTURA (Msnm)	msnm	529	462	1566,40	1580	1000	2000
Beneficio	benefic	748	243	2,13	2	1	4
Tipo De Secado	tsecado	761	230	3,83	4	1	5
Certificación	certific	991	0	6,68	6	1	9
Aspecto Almendra	aspect	988	3	35,88	38	1	46
Humedad CPS (%)	humcsp	979	12	0,12	0,1176	0,075	0,3
Merma (%)	mermap	646	345	0,18	0,1856	0,1476	0,2196
Olor Almendra	coloralm	735	256	3,06	3	1	8
Color Almendra	oloralm	735	256	6,90	7	1	7
Almendra Sana (%)	almendefecp	646	345	0,77	0,7762	0,4704	0,8324
Almendra Defectuosa (%)	almsanap	646	345	0,05	0,0386	0,0016	0,3144
Clases De Defectos	Defec	991	0	7,82	9	1	15
Puntuación final	puntfin	991	0	2,31	2	1	4

Para las variables sin imputación que se muestran en la Tabla 8, las variables altura, merma, almendra sana, almendra defectuosa y sombrío, mostraron el número más alto de valores perdidos en contraste con variables como certificación, aspecto de la almendra, humedad del CPS que presentaron los valores más bajos.

Con este conjunto de variables se realizaron 3 técnicas de imputación de datos Listwise o case deletion CD, Imputación Múltiple - MI, Hot Deck -HD.

En la Tabla 9 se muestran los estadísticos para las variables en estudio. En dicha Tabla se detallan los valores de la media (Me), mediana (Med), desviación estándar (SD), coeficiente de asimetría (As), curtosis (K) y el contraste de normalidad de Kolmogorov-Smirnov-Lilliefors (KSL) bajo las diferentes técnicas de imputación. Estos estadísticos permitieron evaluar e las técnicas de imputación a través de la simulación.

Tabla 9. Variables imputadas bajo las diferentes técnicas.

Variables	Método Imputación	N Válidos	Me	Med	SD	As	k	KSL (=0,05)	
								D	p-valor
Variedad	Imp. Mult	991	10,06	12	3,074	-0,570	-0,500	0,309	0,00
	Hot-Deck	991	10,05	12	3,072	-0,590	-0,416	0,317	0,00
	Listwise	306	10,07	12	3,038	-0,263	-1,078	0,284	0,00
Sombrío	Imp. Mult	991	9,30	5	7,512	0,605	-1,044	0,262	0,00
	Hot-Deck	991	9,54	5	7,694	0,563	-1,129	0,257	0,00
	Listwise	306	8,93	5	7,449	0,670	-0,967	0,267	0,00
Fertilización	Imp. Mult	991	2,93	3	0,346	-5,038	24,338	0,082	0,00
	Hot-Deck	991	2,93	3	0,341	-5,074	24,795	0,537	0,00
	Listwise	306	2,96	3	0,283	-6.604	42,238	0,537	0,00
ALTURA (Msnm)	Imp. Mult	991	1566	1580	136,113	-0,384	0,681	0,533	0,00
	Hot-Deck	991	1561	1570	132,621	-0,283	0,738	0,082	0,00
	Listwise	306	1567	1580	134,542	-0,332	0,513	0,086	0,022
Beneficio	Imp. Mult	991	2,13	2	0,474	3,528	11,164	0,536	0,00
	Hot-Deck	991	2,13	2	0,505	3,284	9,629	0,536	0,00
	Listwise	306	2,16	2	0,572	2,718	6,212	0,524	0,00
	Imp. Mult	991	3,83	4	0,671	-3,848	13,179	0,322	0,00

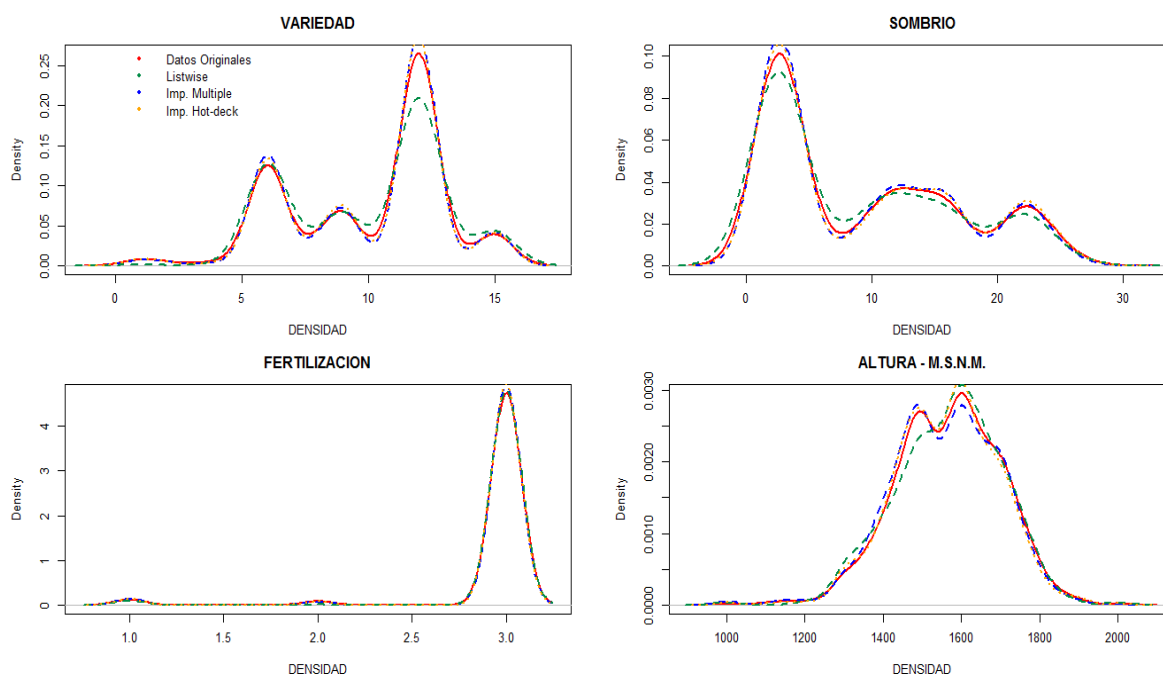
Tipo De Secado	Hot-Deck	991	3,82	4	0,705	-3,624	11,394	0,537	0,00
	Listwise	306	3,78	4	0,769	-3,187	8,500	0,530	0,00
Certificación	Imp. Mult	991	6,68	6	1,779	-0,588	1,411	0,513	0,00
	Hot-Deck	991	6,68	6	1,779	-0,588	1,411	0,322	0,00
	Listwise	306	6.52	6	1,795	-0,651	1,535	0,314	0,00
Aspecto Almendra	Imp. Mult	991	35,88	38	7,256	-3,268	9,602	0.567	0,00
	Hot-Deck	991	35,88	38	7,198	-3,295	9,809	0,514	0,00
	Listwise	306	36,24	38	6,543	-3,742	13,316	0,524	0,00
Humedad Cps	Imp. Mult	991	0,12	0,1176	0,016	2,327	16,772	0,118	0,00
	Hot-Deck	991	0,12	0,12	0,016	2,357	17,055	0,121	0,00
	Listwise	306	0,11	0,11	0,008	-0,496	-0,615	0,086	0.022
Merma	Imp. Mult	991	0,18	0,1856	0,017	-0,177	-0,512	0,057	0.003
	Hot-Deck	991	0,18	0,19	0,012	-0,077	0,326	0,047	0.027
	Listwise	306	0,18	0,18	0,013	0,012	-0,093	0,045	0.0047
Olor Almendra	Imp. Mult	991	3,06	3	0,675	4,613	37,002	0,494	0,00
	Hot-Deck	991	3,06	3	0,711	5,136	35,499	0,512	0,00
	Listwise	306	3,13	3	0,885	4,613	22, 95	0,523	0,00
Color Almendra	Imp. Mult	991	6,90	7	0,587	-6,223	41,130	0,535	0,00
	Hot-Deck	991	6,90	7	0,602	-6,272	41,488	0,536	0,00
	Listwise	306	6,91	7	0,535	0,418	40,055	0,535	0,00
Almendra Sana	Imp. Mult	991	0,77	0,7762	0,043	-1,647	4,496	0,1190	0,00
	Hot-Deck	991	0,77	0,78	0,034	-1.649	4,419	0,125	0,00
	Listwise	306	0,76	0,78	0,045	-1,921	6,478	0,151	0,00
Almendra Defectuosa	Imp. Mult	991	0,05	0,0386	0,036	1,973	6,380	,131	0,00
	Hot-Deck	991	0,05	0,04	0.036	1,834	5,711	0,128	0,00
	Listwise	306	0,05	0,04	0,039	2,062	7,043	0,159	0,00
Clases De Defectos	Imp. Mult	991	7,82	9	2,783	0.664	1,626	0,264	0,00
	Hot-Deck	991	7,82	9	2,783	-0.664	1,626	0,213	0,00
	Listwise	306	8,03	9	3,197	0,913	0,787	0,429	0,00
Puntuación final	Imp. Mult	991	2,31	2	17,143	0,255	-0,529	0,288	0,00
	Hot-Deck	991	2,31	2	17,143	0.255	-0,529	0,288	0,00
	Listwise	306	2,78	2	17,061	0,418	-0,905	0,188	0,00

Uno de los supuestos suele ser evaluar la normalidad de las variables. Como se observa a través de las estadísticas de forma, los coeficientes de asimetría y curtosis

muestran que ninguna de las variables imputadas en este estudio sigue una distribución normal, lo que se corrobora con la prueba KSL que con un nivel de significancia de 0.05, el P-valor de la prueba (la probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido), fue inferior en todas las variables por lo que se rechaza la hipótesis de normalidad.

En la Figura 5, se presentan las distribuciones de probabilidad para las variables en estudio bajo las tres técnicas de imputación. Como se puede observar en los gráficos de densidad, los datos generados por MI y por imputación HD se ajustaron bien a los datos originales lo que no sucedió con la técnica de CD.

Como se muestra gráficamente, ni las variables originales ni las imputadas muestran distribuciones normales, por lo cual se verifica que después de realizar las imputaciones dichas distribuciones no han cambiado su distribución.



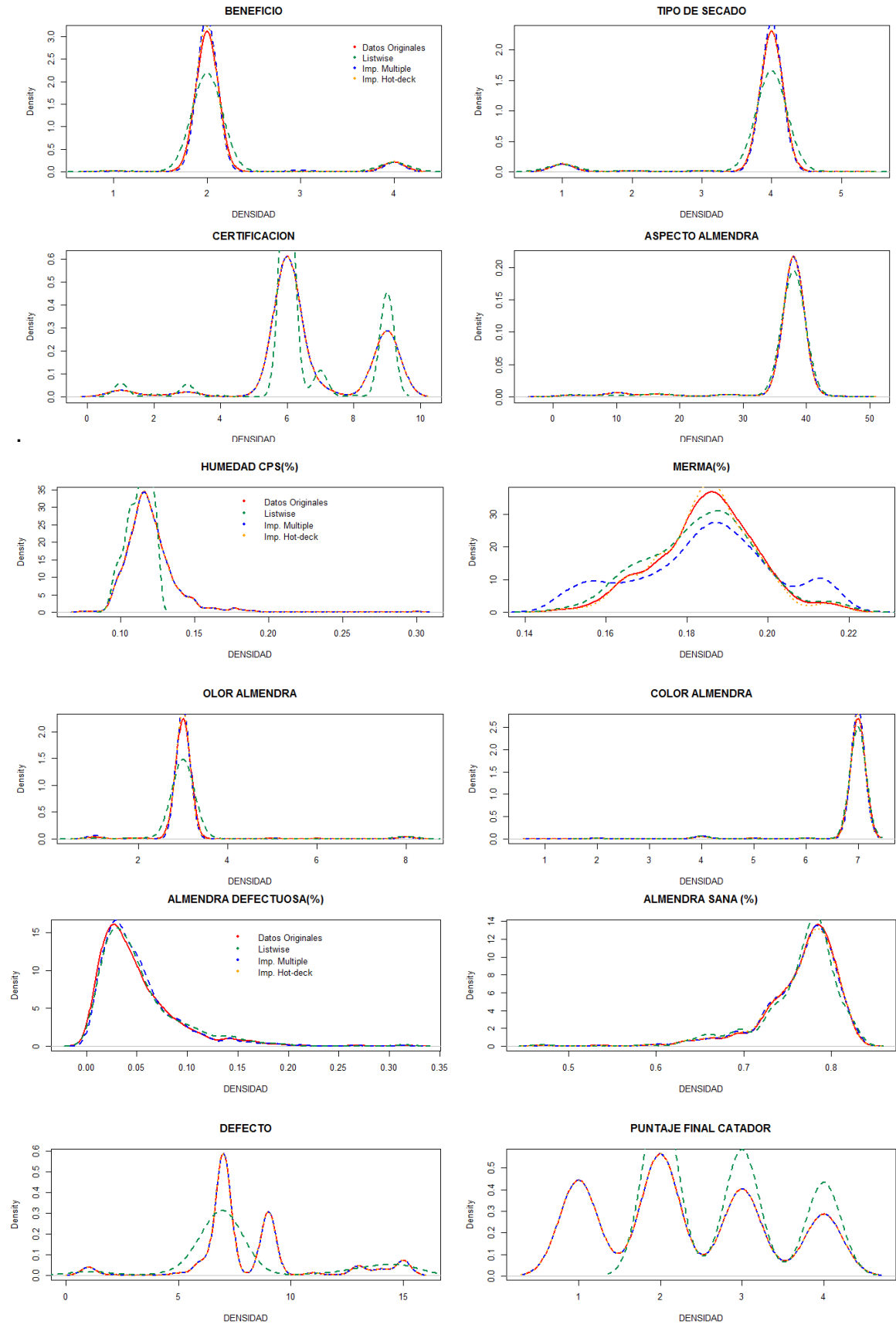


Figura 5. Variables antes y después de los procesos de imputación.

Por último, antes de hacer análisis de los resultados bajo las diferentes técnicas, se calculó el coeficiente de correlación entre las variables lo cual en principio permitiría conocer las relaciones existentes entre las variables, pero debido a que no hay normalidad en las variables, no se pudo calcular coeficiente de correlación de *r Pearson*, Por esta razón se utilizó el coeficiente de correlación no paramétrico de *rs Spearman*, el cual no requiere normalidad y además acepta variables de libre distribución e incluso ordinales.

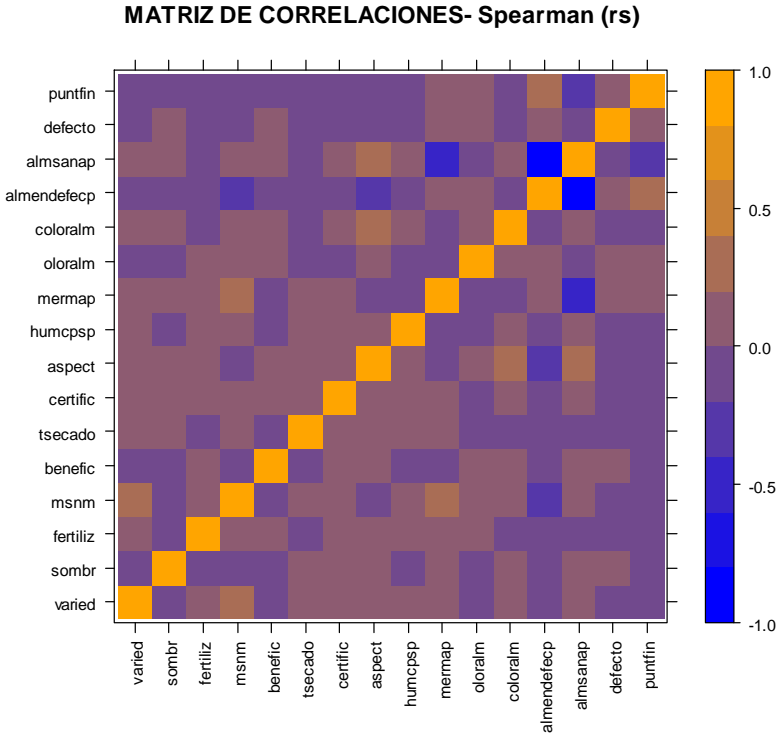


Figura 6. Relación entre las variables iniciales.

Los resultados que se presentan en la Figura 6, muestran los bajos valores del coeficiente de correlación de *rs Spearman*, lo que permite concluir que existe una relación entre las variables aunque sea muy débil, por lo que en este caso, es importante hacer un análisis de la relación entre las variables después de realizar las imputaciones mediante las diferentes técnicas.

4.1.2 ANÁLISIS DE LOS RESULTADOS PARA VALORES FALTANTES

Según los resultados de la Tabla 9, los análisis de los estadísticos y los gráficos construidos, muestran que al comparar las técnicas de imputación, ninguna de ellas afectó significativamente ni la media, ni la mediana, ni siquiera omitiendo valores faltantes por la técnica de CD, una situación obvia debida a que ambas son medidas de tendencia central, por lo cual son muy próximas. Con la técnica HD, en este sentido se hizo un buen trabajo eligiendo donantes que no afectaron estas medidas, debido a que se contó con un volumen de datos relativamente alto y por tanto hubo buena disponibilidad de donantes. Para la técnica de MI, sin embargo, después del cálculo de las estadísticas, cabe preguntar si el proceso de imputación con un número de m imputaciones genera inferencias estadísticas útiles. Para conocer este aspecto se realizó previamente el análisis de la eficiencia relativa de las estimaciones con relación al porcentaje de datos faltantes que se aprecia en la Tabla 10.

Rubín (1987) ilustra que al usar entre 2 y 10 imputaciones se pierde poca eficiencia en la estimación (en relación con un número infinito de imputaciones) cuando la fracción de la información faltante es modesta y genera intervalos de confianza y pruebas de hipótesis cercanas a su cobertura nominal y niveles de significación, respectivamente. Sin embargo, Schafer (1997) muestra que cuando la fracción de la información que falta es muy grande, son necesarias más de 10 imputaciones.

Con relación a lo anterior, fue ineludible realizar el análisis de la eficiencia relativa para saber el número m de imputaciones necesarias bajo el porcentaje de valores faltantes de las variables en estudio, donde se escogió un $m = 12$ imputaciones necesarias para obtener estimaciones confiables, como se muestra en la Tabla 10 y la Figura 7.

Tabla 10. Eficiencia de la estimación a través de imputación múltiple según el número de imputaciones (m) y la fracción de información faltante (λ).

	Variedad	Sombrío	Fertiliz	ALTURA (Msnm)	Benefic	Tipo De Secado	Certifica	Aspecto Almend	Humedad Cps (%)	Merma (%)	Olor Almend	Color Almend	Almend Sana (%)	Almend Defect %	Clases De Defectos	Puntaje final
m	Fraccion de la Informcion Faltante															
	0,432	0,527	0,316	0,873	0,325	0,302	0,000	0,003	0,012	0,534	0,348	0,348	0,534	0,534	0,000	0,000
3	0,874	0,851	0,905	0,775	0,902	0,908	1,000	0,999	0,996	0,849	0,896	0,896	0,849	0,849	1,000	1,000
6	0,933	0,919	0,950	0,873	0,949	0,952	1,000	0,999	0,998	0,918	0,945	0,945	0,918	0,918	1,000	1,000
9	0,954	0,945	0,966	0,912	0,965	0,968	1,000	1,000	0,999	0,944	0,963	0,963	0,944	0,944	1,000	1,000
12	0,965	0,958	0,974	0,932	0,974	0,975	1,000	1,000	0,999	0,957	0,972	0,972	0,957	0,957	1,000	1,000
15	0,972	0,966	0,979	0,945	0,979	0,980	1,000	1,000	0,999	0,966	0,977	0,977	0,966	0,966	1,000	1,000
18	0,977	0,972	0,983	0,954	0,982	0,983	1,000	1,000	0,999	0,971	0,981	0,981	0,971	0,971	1,000	1,000
21	0,980	0,976	0,985	0,960	0,985	0,986	1,000	1,000	0,999	0,975	0,984	0,984	0,975	0,975	1,000	1,000

Según la Tabla 10, Cuando $\lambda = 0.873$, por ejemplo, una estimación basada en $m = 3$ imputaciones, tenderá a tener un error estándar sólo en un $\sqrt{1 + \left(\frac{0,873}{3}\right)} = 1.136$ veces más grande que el estimado con $m = \infty$ ó con $\lambda = 0.012$, una estimación basada en $m = 21$ imputaciones tienden a tener un error estándar de $\sqrt{1 + \left(\frac{0,012}{21}\right)} = 1.00$ veces más grande que el estimado con $m = \infty$.

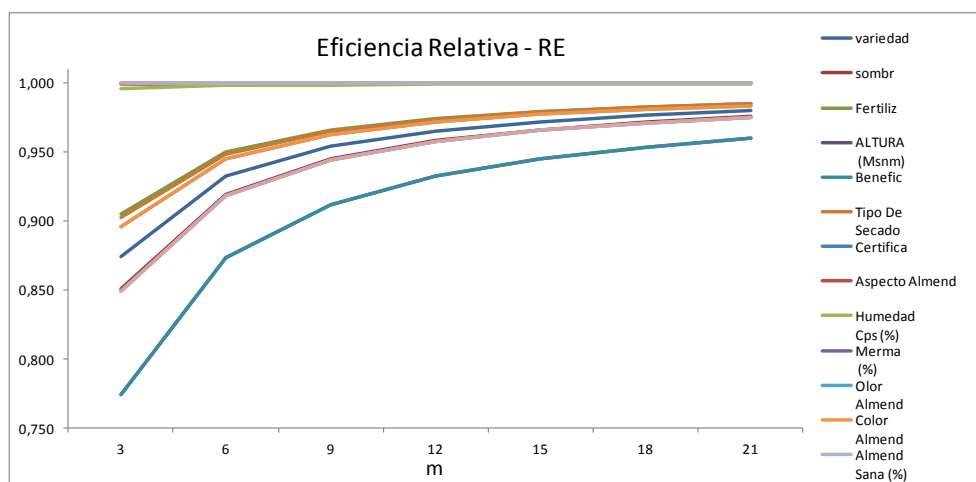
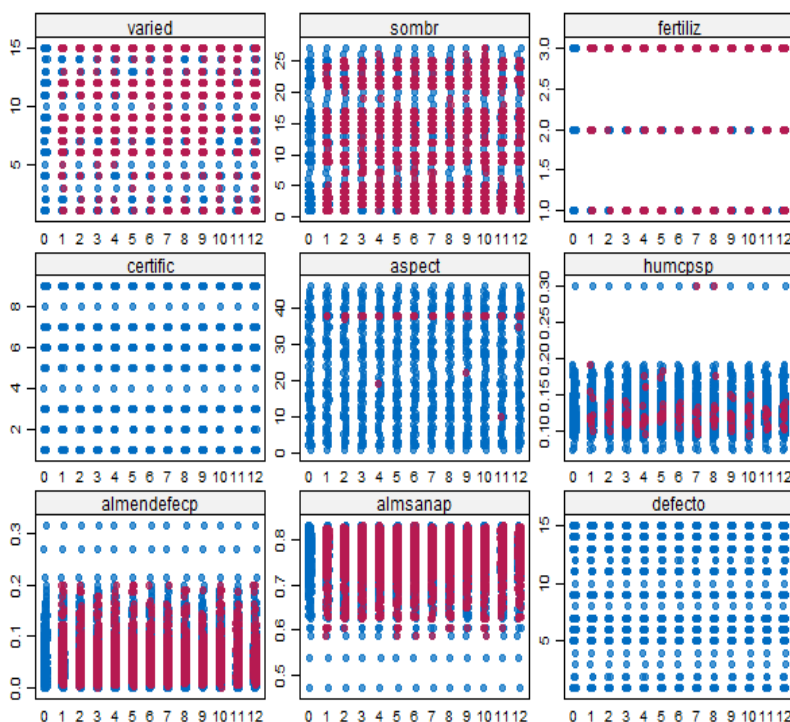


Figura 7. Eficiencia Relativa de la estimación a través de imputación múltiple según el número de imputaciones (m) y la fracción de información faltante (λ).

Ahora, si se observa la Figura 7, se puede ver que a medida que el número de

imputaciones aumenta, la eficiencia relativa también lo hace. Por esta razón se escogió un tamaño m que tuviera una buena eficiencia sin exagerar el número de imputaciones. Con la técnica de imputación múltiple también se lograron estimar los valores que no afectaran ni la media, ni la mediana. Es por ello que para realizar la segunda parte del estudio y basándose en lo descrito por Rubín (1987, 1996), se consideró que este método es el más adecuado para tratar los problemas más complejos cuando hay un número considerable de datos en más de una variable. Por otro lado, como también se requiere mantener la distribución de frecuencia de la variable y las asociaciones entre las distintas variables, se deben emplear métodos más elaborados aplicando imputación de todas las variables faltantes del registro como la MI. Goicoechea (2002),

Para tener una idea más clara de lo acontecido con el proceso MI, en el Figura 8 que se muestra a continuación se representa la distribución de las variables como puntos individuales, los puntos azules son los valores observados y los puntos rojos son los valores imputados.



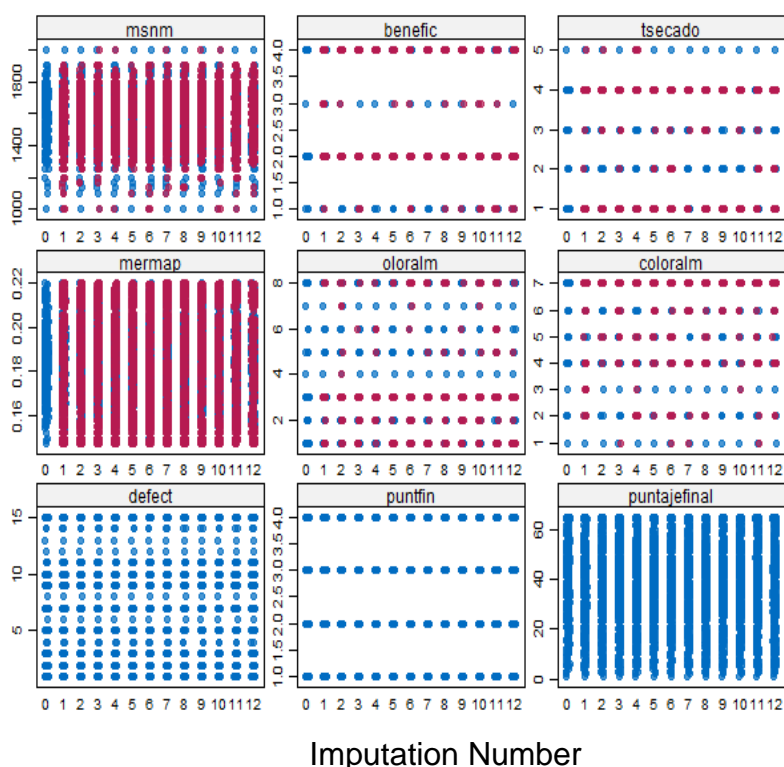


Figura 8. Variables antes y después del proceso de imputación múltiple MI.

En la Figura 8, también se observa que para las variables en estudio como defecto y puntuación final y certificación, solo contiene valores azules, es decir que en los datos originales, estas variables tenían todos sus datos completos, sin embargo, en variables como merma, sombrío, almendra sana y altura, se puede observar que los valores más notorios fueron los imputados (color rojo). A pesar de esto, se puede observar que el método MI respondió bastante bien, ya que los puntos rojos siguen los puntos azules muy razonablemente, incluyendo las brechas en las distribuciones.

4.1.3 ANÁLISIS FACTORIAL BAYESIANO

Una vez que se ejecutó el proceso MI con $m = 12$ y se obtuvo una base de datos completa, se realizó el análisis factorial bayesiano, empleando el paquete MCMCpack de R, a través de la función MCMCmixfactanal. Como resultado se obtuvo una cadena de Markov con algoritmo Monte Carlo – MCMC que contiene una muestra de la densidad a posterior con los datos mixtos del modelo de análisis factorial, es decir datos tanto continuos como ordinales.

El ajuste del modelo se llevó a cabo mediante lo descrito previamente, las primeras 5000 exploraciones de las MCMC fueron descartadas como burn-in. Los resúmenes posteriori de los parámetros del modelo (factor de cargas (λ_2), varianzas del Error (ψ_{jj}) y factor latente de puntuaciones (ϕ_i)), fueron basados en una muestra a posterior de tamaño 5000, formado mediante la ejecución de la cadena por un período adicional de 500000 exploraciones las cuales se fueron almacenando cada 10 exploraciones. La cadena mixta generada, mostró un diagnóstico estándar razonablemente bueno.

```
Posteriori<- MCMCmixfactanal((~varied+sombr+fertiliz+msnm+benefic+
  tsecado+aspect+ humcsp+mermap+oloralm+coloralm+
  almendefecp+almsanap+puntfin), factors=1,
  data=Datos.Imputados,burnin=5000,mcmc=500000,thin=10,
  store.scores=TRUE,tune=0.25,store.lambda=TRUE)
```

Después, se verificó mediante la prueba de Heidelberger and Welch (1992) que la muestra de la cadena Markov generada, alcanzara la distribución estacionaria. Las variables se pueden observar en la función del modelo de ajuste. Solo se eligió estimar un factor (Índice de calidad), el intervalo de precisión fue de 0.25, tanto para almacenar la muestra del factor de cargas (Λ), las varianzas del Error (Ψ) y el factor latente de puntuaciones (Φ) de la distribución posteriori. Básicamente lo que se pretendió, fue aventurarse en la estructura de los datos de un modelo construido con anticipación, para tratar de explicar las correlaciones entre el conjunto de variables observadas a través de un conjunto más reducido de factores latentes.

Resúmenes de la densidad posteriori

A continuación se muestran los resúmenes del factor de cargas (Λ) y la varianzas del Error (Ψ) de la densidad posterior que proporciona información acerca de los patrones globales de la asociación entre las variables observadas del estudio y las estimaciones de las puntuaciones de los factores latentes Φ que finalmente son las

que nos llevan al índice de calidad.

Factor de cargas (Λ) y Varianzas del Error (Ψ)

El factor loading o factor de cargas (Λ) permite observar las conexiones entre las variables y el factor, es decir, indica la fuerza de la dependencia de cada variable observada en cada factor.

En la Tabla 11, se muestran las variables que constituyen las propiedades organolépticas y agronómicas obtenidas en las diferentes pruebas de tasa, la estimación de la variable almendra defectuosa en el factor de cargas es de -1.00, lo que indica que presenta una asociación negativa muy alta entre esta variable y el factor, en contraste con las variables como merma y almendra sana que presentaron una asociación positiva de 0.5529 y 0.9269 respectivamente, el resto de las variables mostraron valores muy bajos.

Tabla 11. Factor de cargas (λ_2) y varianza del Error (ψ_{jj}).

Variables	(λ_2)	(ψ_{jj})
Variedad	-0,0574	0,9975
Sombrío	-0,0625	0,9983
Fertilización	-0,0013	1,0015
Altura (Msnm)	-0,1794	0,9711
Beneficio	0,0253	1,0005
Tipo De Secado	0,0433	1,0008
Aspecto Almendra	-0,1019	0,9910
Humedad CPS	0,0087	1,0015
Merma	0,5529	0,6998
Olor Almendra	0,0553	0,9990
Color Almendra	0,0782	0,9953
Almendra Sana	0,9268	0,1509
Almendra Defectuosa	-1,0052	0,0007
Puntuación final	0,1312	0,9855

Con las varianzas del error se puede tener una idea de la varianza de la variable que

puede ser explicada por el factor latente, es decir que para la almendra defectuosa, la varianza del error estimado se calcula en 0.0007, lo que implica que casi toda la variabilidad de la almendra defectuosa puede ser explicada solo por el factor latente, débilmente también se podrían tener en cuenta la merma, olor el color y la almendra sana.

Factor latente de puntuaciones (ϕ_i)

El factor latente de puntuaciones es una variable no observada que se compone de numerosas variables (observables), esta variable permite explicar la variabilidad común que tienen todas ellas a través de las puntuaciones, es decir, que es una forma de consolidar numerosa información en una sola variable. Por ello, la determinación de este factor latente fue el interés principal de este estudio, con él se obtuvo el índice de calidad que se puede resumir en la Tabla 12 para una mejor interpretación.

Tabla 12. Frecuencias para el Índice de calidad – IC (ϕ_i).

Intervalo		Media	Frecuencia Absoluta	Frecuencia relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada	Valor IC Café
Inf.	Sup.						
-2	-1	-1,5	88	0,089	88	0,089	Mala calidad
-1	0	-0,5	519	0,524	607	0,613	
0	1	0,5	265	0,267	872	0,88	media
1	2	1,5	67	0,068	939	0,948	Buena
2	3	2,5	34	0,034	973	0,982	
3	4	3,5	15	0,015	988	0,997	excelente
4	5	4,5	1	0,001	989	0,998	
5	6	5,5	1	0,001	990	0,999	
6	7	6,5	1	0,001	991	1	

En la Tabla 12, se puede ver que el 61,3 % de los datos presenta puntuaciones entre [-2,0] lo que se ha valorado como un índice de calidad malo, el 26.7% de las puntuaciones entre [0,1] catalogado como calidad media, el 10,2% entre [1,3] como calidad buena y solo el 1.8% de las puntuaciones presentaron una excelente calidad

con valores superiores a 3.

En las Figuras 9 -14 que se presentan a continuación se puede contrastar el índice de calidad con otras variables de interés.

Por Municipio

Como se ve en la Tabla 13 y Figura 9, existe una diferencia bien marcada entre las calidades del café, con respecto a las zonas, por ejemplo la zona norte del Departamento muestra una calidad excelente, la zona Centro una calidad buena y la zona sur una calidad media. Eso era esperado dado que la zona norte (El Águila, Alcalá, el Dovio, Roldanillo, Bolívar, Trujillo, Bugalagrande y Sevilla) y centro del departamento (Riofrío, Tulúa, San Pedro, Buga, Guacarí, Ginebra y Buga) posee el entorno típico de la zona cafetera, es decir, que en cuanto a los aspectos agronómicos ideales del cultivo como suelos muy fértiles, temperaturas entre los 17 y los 23 grados centígrados, precipitaciones cercana a los 2.000 milímetros anuales y altura entre los 1.200 y 1.800 metros sobre el nivel del mar muestra que se pueden producir cafés de excelente calidad.

En cuando a la zona Sur del departamento, se muestra que también es posible producir un café con una calidad sobresaliente a alturas marginales y con niveles o frecuencia de precipitación diferentes. Esta zona pues, podría ser de especial interés, debido el alto potencial que presentan municipios como Pradera, Palmira y El Cerrito, además de municipios como Dagua, florida, la cumbre, Yótoco, Restrepo y Yumbo para producir un café con muy buena calidad. Finalmente, cabe anotar que este mapa es una primera aproximación a la zonificación por pruebas de taza y en ningún caso es una verdad absoluta, sin embargo refleja una realidad que se puede corroborar con otros estudios realizados por Cenicafé.

Tabla 13. Municipios del Valle del Cauca que forman parte del estudio.

#	Municipio	#	Municipio	#	Municipio	#	Municipio
1	Yumbo	10	Riofrío	19	Florida	28	Buga
2	Yótoco	11	Restrepo	20	El Dovio	29	Bolívar
3	Ulloa	12	Pradera	21	El Cerrito	30	Argelia
4	Tulúa	13	Palmira	22	El Cairo	31	Ansermanuevo
5	Trujillo	14	Obando	23	El Águila	32	Andalucía
6	Toro	15	La Cumbre	24	Dagua	33	Alcalá
7	Sevilla	16	Jamundí	25	Cali		
8	San Pedro	17	Guacarí	26	Caicedonia		
9	Roldanillo	18	Ginebra	27	Bugalagrande		

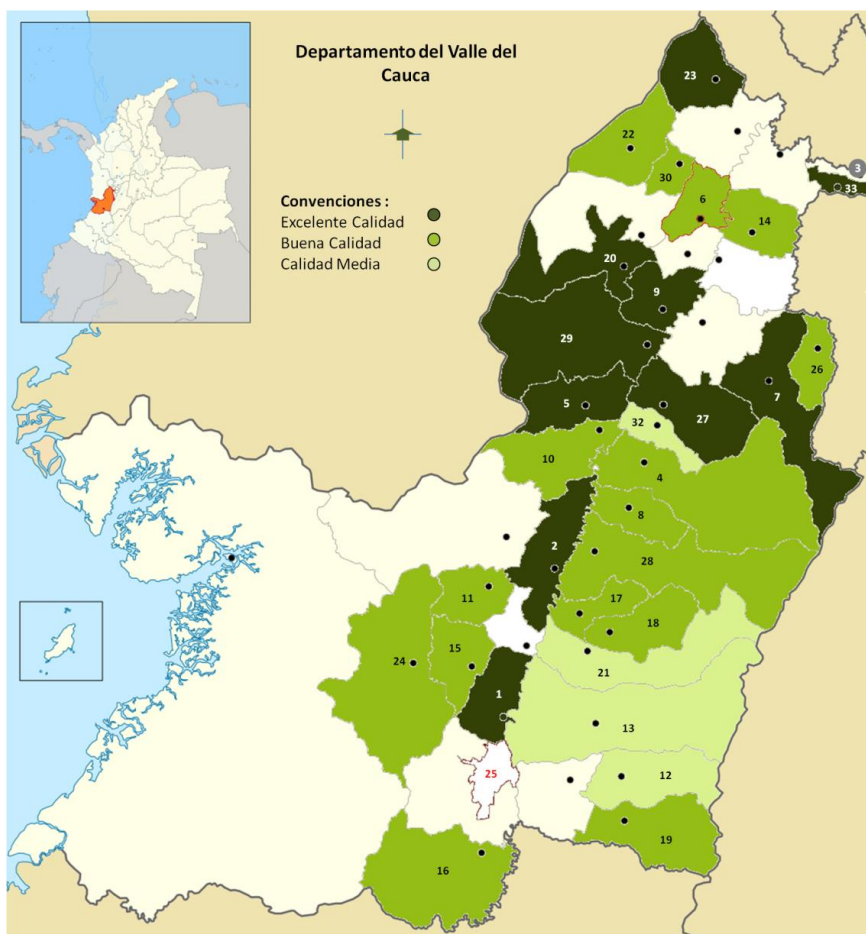


Figura 9. Mapa del Departamento del Valle del Cauca que presenta una media, buena y excelente calidad.

Por Variedad

En el Figura 10A se observa que la variedad predominante en el estudio es la Caturra con el 49.3%, seguido Colombia 22.2% y la Castillo con 7.5%, además se muestra la combinación de estas variedades dentro del cultivo como la Caturra y la Colombia con un 11.8% y Caturra y Típica 3.5%.

#	Variedad	#	Variedad
1	Típica	9	Caturra y Colombia
2	San Bernardo	10	Caturra y Catimor
3	Costarica	11	Caturra y Castillo
4	Colombia, Caturra y Típica	12	Caturra
5	Colombia y Típica	13	Catimor
6	Colombia	14	Castillo y Colombia
7	Caturra, Colombia y tabi	15	Castillo
8	Caturra y Típica		

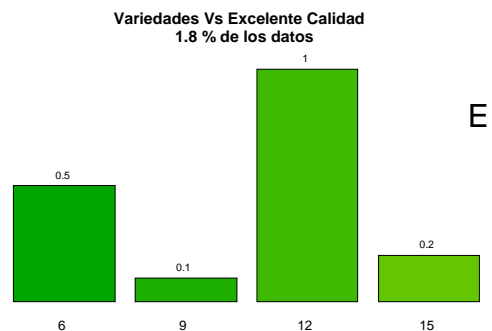
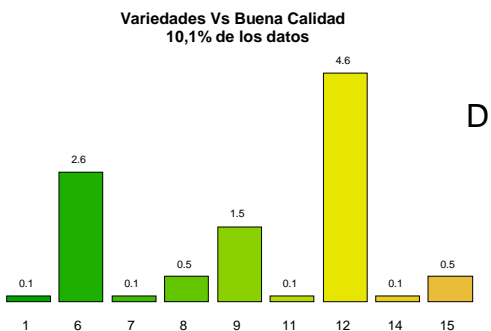
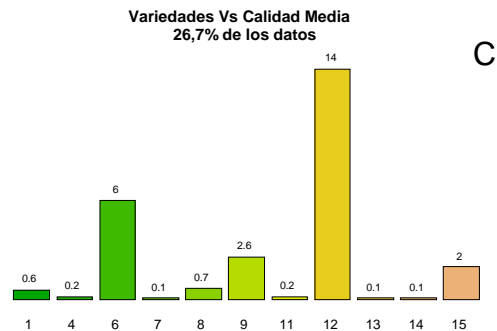
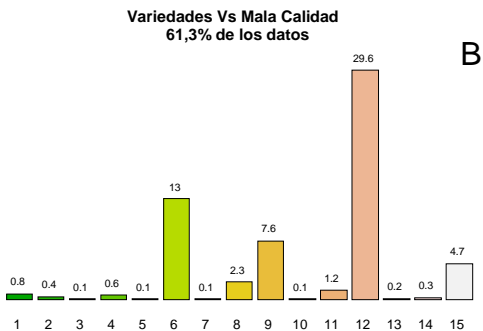
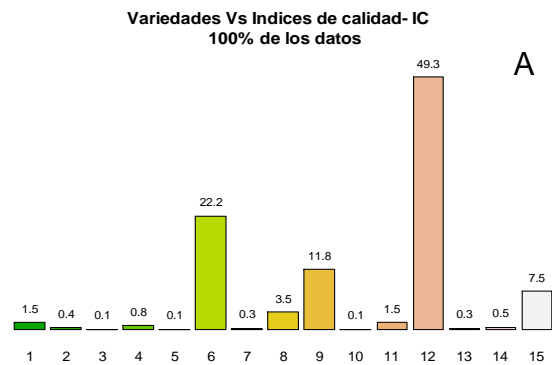


Figura 10. Variedades de café según el Índice de calidad.

Ahora si se observa la Figura 10E, que muestra una excelente calidad según el índice hallado, se puede ver que siguen estando presentes estas cuatro

variedades, así como 10B, 10C y 10D. Estos valores concuerdan con estudios realizados por Cenicafe donde la variedad Colombia, Caturra, Típica y Castillo se destacan por su calidad en taza con relación a otras variedades (Puerta 1998, 2000b: Alvarado & Puerta, 2002).

Por Certificación

La certificación del café son signos que aseguran al consumidor que el café cumple con unos estándares de calidad previamente definidos en normas legales, técnicas o reglamentos.

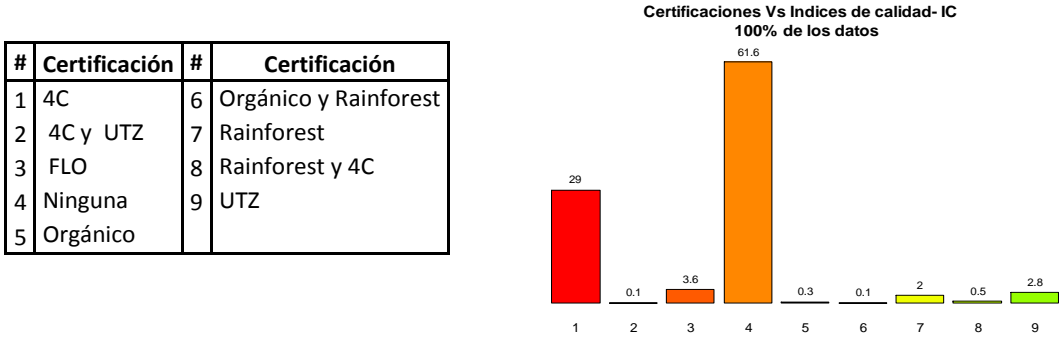


Figura 11. Certificación de café según el Índice de calidad.

Según la Figura 11, el 61.6% de las pruebas de taza realizadas a los café no tenía ningún tipo de certificación, el 29% presentó la certificación 4C(cafés procedentes de procesos productivos con buenas prácticas sociales, económicas y ambientales), el 3.6% la certificación de Comercio Justo - FLO (cafés que provienen de un proceso productivo donde se protege el medio ambiente lo que permite recibir un precio de venta mayor y más estable), el 2% la certificación Rainforest (el manejo sostenible de los cafetales y las fincas es orientado a la protección del medio ambiente) y el 2.8% la certificación UTZ (cafés producido bajo estrictas normas de calidad).

Por Puntuación Sensorial

La Figura 12, muestra la puntuación sensorial del café. Aquí, el 33.3% de las tazas presentaron muy buen perfil, el otro gran porcentaje lo contienen las muestras que

no son Aptas 26.1% lo que quiere decir que presentaron una humedad muy alta y por lo tanto, afectan directamente la calidad, el 23.7% de las pruebas de taza poseen algún defecto que se mostraran más adelante y el 16.9% restante mostraron un buen puntuación sensorial.

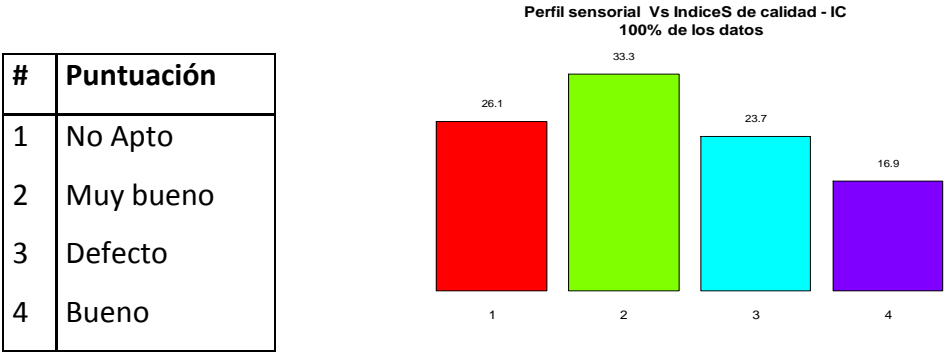


Figura 12. Puntuación sensorial del café según el Índice de calidad.

Por Defecto

La Figura 13, permite ver los defectos presentes en las pruebas de taza. Cabe resaltar que el 49.8% de las pruebas tuvieron una taza limpia (sin defectos), y el 26.1% como se mostró anteriormente no fueron aptas para realizar las pruebas.

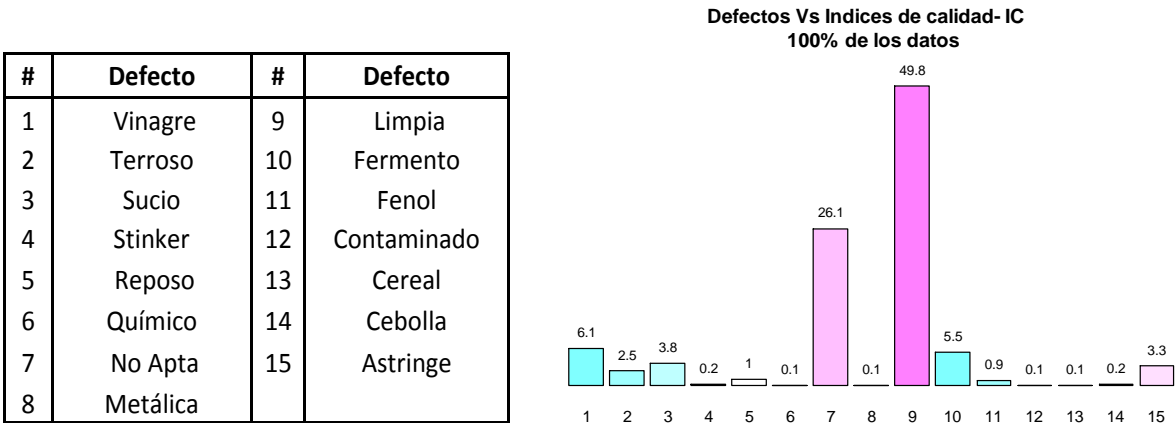


Figura 13. Defectos presentes en las pruebas de taza contrastados con el índice de calidad.

Los defectos más predominantes en las restantes muestras fueron vinagre 6.1%

(causado básicamente por retrasos entre la recolección y el despulpado, fermentaciones demasiado prolongadas, uso de aguas sucias y almacenamiento húmedo del café), fermento 5.5% (se favorece el mal despulpado y por la recolección de granos sobremaduros y verdes), sucio 3.8% (mal lavado del café), astringente 3.3% (ocasionado por granos inmaduros), terroso 2.5% (se da en cafés que están en contacto con el suelo o agua sucia), reposo 1% (ocurre cuando los granos se almacenan durante mucho tiempo o se guardan en condiciones inadecuadas y en sitios muy calientes) y fenol 0.9% (se produce por café pergamino húmedo amontonado antes de secado).

4.1.4 USO DE LA IMPUTACIÓN MÚLTIPLE Y ANÁLISIS FACTORIAL BAYESIANO

El uso de la imputación de datos es una herramienta muy útil cuando se presentan datos faltantes, sin embargo, el uso de las técnicas para imputar requieren de atención, ya que su uso inadecuado podría generar datos imputados inconsistentes produciendo una base de datos no confiable y por lo tanto invalidar el estudio. Por el contrario, si se utiliza correctamente, se puede obtener un conjunto de datos completo, se puede reducir el sesgo debido a la ausencia de información y obtener resultados consistentes. Por el momento no existen criterios para decidir el porcentaje máximo de valores faltantes que deben aceptarse, no obstante, todo depende de la información que posea la base de datos y la aplicación que se vaya a realizar, ya que siempre habrán estudios más sensibles que otros a los valores faltantes.

En este estudio se utilizaron tres técnicas de imputación de datos Listwise o case deletion CD, Imputación Múltiple – MI y la Hot Deck – HD, de todas ellas, hasta el momento la más confiable es la MI, dada las bondades de la técnica aportada por investigaciones previas de Azur, et al. (2011), Van & Groothuis (2011) Raghunathan et al. (2001), Schafer & Graham (2002), Graham (2009), Stuart et al. (2009), Allison (2002). Otero (2011), Goicochea (2002) , Rubín & Schenker (1986) ,Van (2007) y Amón, (2010), donde menciona que la MI es una técnica que incrementa la eficiencia de los estimadores, genera inferencias validas, permite calcular el error

estándar de los estimadores, utiliza procedimientos estadísticos robustos, genera distintas opciones de datos imputados y los combina de una forma adecuada. Por otro lado señalan que la CD, genera una pérdida substancial de la información en bases de datos, altera el tamaño de la muestra y el número de variables, y además supone que las observaciones de los valores que faltan no son importantes y que pueden ser ignorados por los estimadores de los parámetros, generando así, estimadores sesgados que en resumen podrían invalidar las conclusiones del estudio. La técnica HD tiene la limitante que cuando se presenta una gran cantidad de datos faltantes se obtiene muchas veces mismo valor duplicado, lo que distorsiona la relación con el resto de las variables y como consecuencia genera un mecanismo que carece de probabilidad lo que imposibilita calcular su confianza y además puede proporcionar estimaciones sesgadas de la media (Carter, 2006; Medina & Galván, 2007; Magnani, 2004).

Con el método de imputación múltiple seleccionado, se generó una base de datos completa que dio mejores aproximaciones a los valores verdaderos, mostró una buena eficiencia, logró estimar valores que no afectarían ni la media, ni la mediana, mantuvo la distribución de frecuencia de la variable, marginal o conjunta y mantuvo las asociaciones entre las distintas variables.

El análisis factorial es uno de los métodos más utilizados del análisis multivariante para explorar la estructura de covarianzas entre un conjunto de variables aleatorias observadas mediante la construcción de un número menor de variables aleatorias llamados factores comunes. Sin embargo, muy a menudo surge que las estimaciones por máxima verosimilitud de las varianzas únicas llegan a ser cero o negativas siendo inadecuadas para interpretar el modelo estimado. Con el fin de prevenir este tipo de situaciones, se utilizó una aproximación bayesiana mediante la especificación de la distribución a priori de los parámetros para un modelo de factor de análisis. En este caso se emplearon las distribuciones a priori propuestas por Quinn (2004), que se utilizan cuando las respuestas son continuas, ordinales, dicotómicas y politómicas, demostrando ser eficaz numéricamente (Hirose, et al. 2008; Conti, et al. 2014).

De este análisis, se obtuvo un índice de calidad del café que guarda mucha relación con estudios previos realizados por Arcila et al. (2007); Puerta 1998, 1999, 2000a) donde se relaciona con características agronómicas del cultivo y con características organolépticas para determinar la calidad de la bebida.

4.2 CALIDAD DEL CAFÉ ESPRESSO A PARTIR DE LAS CUALIDADES FISICOQUÍMICAS

Los resultados que se muestran a continuación hacen énfasis en los análisis por tipo de café, en este caso, A (Robusta Natural), B (Arábica Natural) y C (Arábica Lavado), así como también en los perfiles o curvas de temperatura utilizados ascendente (88-93 °C), plana (90-90 °C) y descendente (93-88 °C).

4.2.1 GRANULOMETRÍA

En la Tabla 14 y Figura 14, se muestra la clasificación de la molienda de los tipos de cafés con los diferentes perfiles de temperatura para obtener un volumen de CE en un tiempo de erogación establecido.

Tabla 14. Tamaño de partícula y clasificación de la molienda.

CAFÉ	TEMPERATURA DE LA MAQUINA °C	CLASIFICACIÓN DE LA MOLIENDA (%)		
		Gruesa >500 µm	Media 500-250 µm	Fina < 150 µm
A	90-90	45,16	39,57	15,26
	88-93	44,64	41,87	13,48
	93-88	48,86	44,2	6,93
B	90-90	38,87	49,61	11,52
	88-93	42,18	45,71	12,1
	93-88	48,24	42,55	9,19
C	90-90	32,71	56,69	10,58
	88-93	38,04	50,54	11,42
	93-88	46,46	45,18	8,35

Si se analiza la granulometría por el tipo de café, se encuentra que para el café A (Robusta Natural), el mayor porcentaje de partículas gruesas ($\geq 500 \mu\text{m}$) se presentó en la temperatura de extracción de A (93-88 °C) con un 48.86%, así como el contenido de partículas medias (500 - 250 μm) con un 44.2%. Por el contrario, para la temperatura 90-90 °C el porcentaje de partículas finas ($\leq 150 \mu\text{m}$) fue mayor con un 15.25%. Para el Café tipo B (Arábica Natural), el mayor porcentaje de partículas gruesas se encontró en la misma temperatura que el café tipo A (93-88 °C) con 48.24%, el mayor contenido de las partículas medias se obtuvo en la temperatura 90-90 °C con un 49.61% y las partículas finas en la temperatura 88-93 °C con 12.1%.

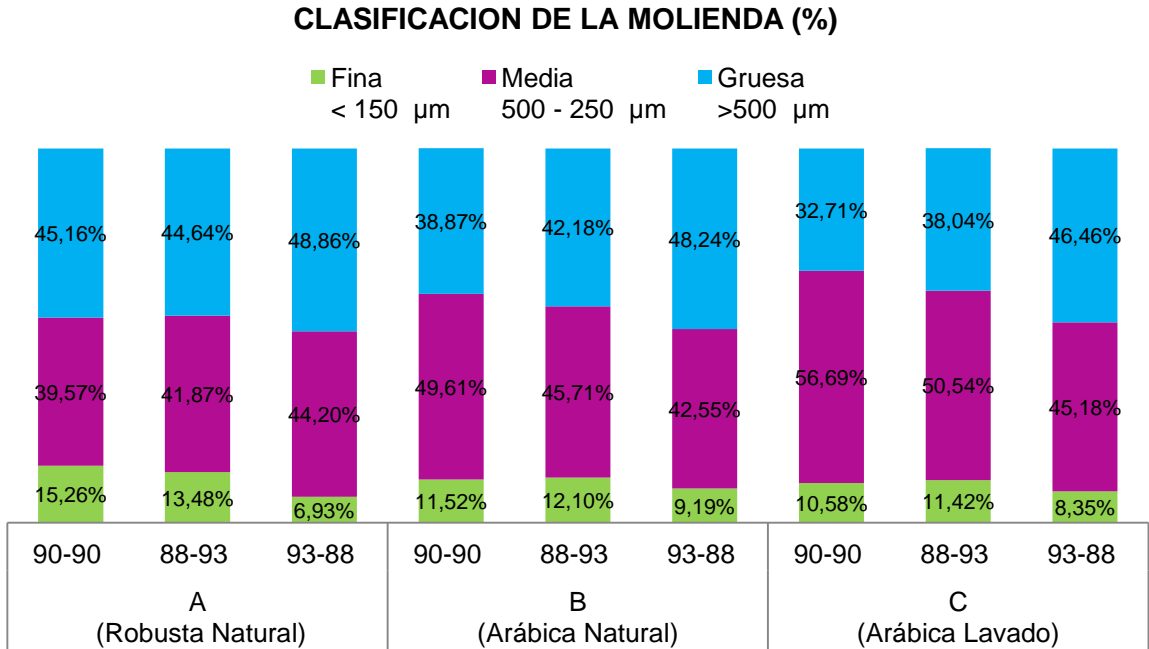


Figura 14. Distribución del tamaño de partículas según El tipo de Café y la temperatura.

Para el café C (Arábica Lavado), el porcentaje más alto de partículas gruesas fue en la temperatura 93-88 °C con 46.46%, las partículas medias en la temperatura 90-90 °C con 56.69% y las partículas finas en la temperatura 88-93 °C con 11.42%.

Ahora, si se analiza la distribución de las partículas, el porcentaje más alto de

partículas gruesas se obtuvo en el café A (93-88 °C) con 48.86% y el menor en C (90-90 °C) con 32.71%, en las partículas medias, el café C (90-90 °C) con 56.39% fue el mayor y el menor fue 39.57% en el café A (90-90 °C) y en las partículas finas el mayor contenido fue para el café A (90-90 °C) con 15.26% y el menor se presentó en el café A (93-88 °C) con 6.93%

El objetivo de la molienda según Cappuccio et al. (2001), es de obtener una distribución de partículas adecuadas para ofrecer un fluido adecuado. Una distribución de tamaño de partículas gruesas que permite el flujo correcto a través de la torta y las partículas medias a finas forma una alta superficie específica, que permiten la extracción de gran cantidad de material soluble y emulsionable. Paralelamente, la molienda destruye la estructura celular del café tostado y esto se convierte en una notable liberación de dióxido de carbono con consecuencias obvias en la crema y en los aromas.

4.2.2 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES

En la Tablas 15 Y 16, se pueden observar las estadísticas descriptivas de las variables químicas y físicas teniendo en cuenta los dos factores, el café y la temperatura.

Como se observa en la Tabla 15 el índice de la crema en los cafés A (88-93 °C) y A (90-90 °C) con 33.24 y 33.47 respectivamente, fueron superiores a los demás cafés en donde los índices más bajos fueron para C (93-88 °C) con 16.82 y C (90-90 °C) con 17.78. La densidad, en general para todos los cafés fue muy uniforme contemplado valores entre 1.02 y 1.03 g/mL. Para la turbidez, las mayores Unidades Nefelométricas de turbidez (NTU) las presentaron los cafés C (90-90 °C) con 2492.17, el A (90-90 °C) con 1844.58 y el C (93-88 °C) con 1759.17, en contraste con los cafés B (88-93 °C) y el A (88-93 °C) que obtuvieron valores de 943.06 y 973.06 NTU respectivamente. El contenido de sólidos totales más altos se presentaron en los cafés B (88-93 °C) con 68.9228 g/L, A (90-90 °C) con 66.87 g/L y A (88-93 °C) con 63.83 g/L y los más bajos en los cafés C (88-93 °C) con 53.6833 g/L y C (93-88 °C) con 54.5211 g/L. Ahora, si se observa los sólidos filtrados, se

puede decir que los cafés que mostraron mayor cantidad de ellos fueron los cafés B (88-93 °C) con 5.9715 g/L y C (90-90 °C) con 4.8089 g/L, los contenidos más bajos fueron para A (88-93 °C) con 2.0461 g/L y A (93-88 °C) con 2.2511 g/L. Finalmente para la variable viscosidad expresada en $\text{mN/m}^2\text{s}^{-1}$ los valores más altos fueron para los cafés C (88-93 °C) con 1.4250 y A (88-93 °C) con 1.3626 respectivamente, por el contrario el valor más bajo se obtuvo en el café tipo A (93-88 °C) con 1.2522.

Tabla 15. Estadísticas descriptivas de las variables físicas (n = 18).

Café	Temperatura °C	Estad.	Índice de la crema	Densidad g/mL	Turbidez NTU	Sólidos Totales g/L	Sólidos Filtrados g/L	Viscosidad $\text{mN/m}^2\text{s}^{-1}$
Tipo A (Robusta Natural)	88-93	Media	33,24	1,03	973,06	63,8333	2,0461	1,3626
		SD	2,56	0,00	159,69	3,2338	0,4802	0,0180
		CV	0,08	0,00	0,16	0,0507	0,2347	0,0132
	90-90	Media	33,47	1,03	1844,58	66,8722	3,0528	1,3013
		SD	1,98	0,01	78,56	2,3232	0,7234	0,0138
		CV	0,06	0,01	0,04	0,0347	0,2370	0,0106
	93-88	Media	23,21	1,03	1110,97	56,2656	2,2511	1,2522
		SD	1,39	0,01	105,07	2,8208	0,4840	0,0201
		CV	0,06	0,01	0,09	0,0501	0,2150	0,0160
Tipo B (Arábica Natural)	88-93	Media	22,87	1,03	943,06	68,9228	5,9717	1,2652
		SD	2,05	0,01	276,31	7,4861	1,2416	0,0176
		CV	0,09	0,01	0,29	0,1086	0,2079	0,0139
	90-90	Media	23,78	1,03	1282,50	59,8833	3,3111	1,2928
		SD	2,80	0,00	245,99	3,1187	0,5527	0,0112
		CV	0,12	0,00	0,19	0,0521	0,1669	0,0086
	93-88	Media	21,94	1,02	1291,08	60,7556	3,7811	1,2690
		SD	1,47	0,02	55,90	3,4648	0,8244	0,0247
		CV	0,07	0,02	0,04	0,0570	0,2180	0,0195
Tipo C (Robusta Lavado)	88-93	Media	19,63	1,03	1212,22	53,6833	3,5694	1,4250
		SD	2,05	0,00	183,54	2,3483	0,3894	0,0602
		CV	0,10	0,00	0,15	0,0437	0,1091	0,0423
	90-90	Media	17,78	1,02	2492,17	63,7594	4,8089	1,2630
		SD	2,53	0,01	356,87	2,2029	0,9033	0,0077
		CV	0,14	0,01	0,14	0,0346	0,1878	0,0061
	93-88	Media	16,82	1,02	1759,17	54,5211	2,7733	1,2891
		SD	1,78	0,00	136,33	1,8091	0,3774	0,0049
		CV	0,11	0,00	0,08	0,0332	0,1361	0,0038

Tabla 16. Estadísticas descriptivas de las variables químicas (n = 18).

Café	Temperatura °C	Estad.	pH	Color	Polifenoles g/L	Cafeína g/L	Lípidos g/ L	Ácido Clorogénico g/ L
Tipo A (Robusta Natural)	88-93	Media	5,01	5,7917	12,1632	2,8094	0,0603	0,4294
		SD	0,8	0,7789	1,9961	0,5200	0,0052	0,0214
		CV	0,2	0,1345	0,1641	0,1851	0,0857	0,0497
	90-90	Media	5,58	5,1025	8,5905	2,4450	0,0871	0,3537
		SD	0,0	0,7015	1,2084	0,3791	0,0093	0,0090
		CV	0,0	0,1375	0,1407	0,1551	0,1068	0,0255
	93-88	Media	5,71	3,7363	7,3693	1,1946	0,0326	0,3131
		SD	0,0	0,3770	0,7554	0,2727	0,0101	0,0080
		CV	0,0	0,1009	0,1025	0,2283	0,3088	0,0256
Tipo B (Arábica Natural)	88-93	Media	5,28	4,9285	11,7882	1,7399	0,0770	0,5099
		SD	0,0	0,8835	3,4538	0,2865	0,0085	0,0332
		CV	0,0	0,1793	0,2930	0,1647	0,1102	0,0652
	90-90	Media	5,74	5,0907	7,3711	2,5591	0,1084	0,4925
		SD	0,0	0,8130	0,9481	0,4471	0,0165	0,0105
		CV	0,0	0,1597	0,1286	0,1747	0,1519	0,0213
	93-88	Media	5,17	2,0058	7,9191	2,1396	0,0635	0,4733
		SD	0,0	0,1785	1,1557	0,2140	0,0062	0,0294
		CV	0,0	0,0890	0,1459	0,1000	0,0980	0,0622
Tipo C (Arábica Lavado)	88-93	Media	5,30	4,6639	15,1528	1,5981	0,0664	0,5553
		SD	0,0	0,4716	2,2942	0,4503	0,0054	0,0265
		CV	0,0	0,1011	0,1514	0,2818	0,0819	0,0477
	90-90	Media	5,34	2,2008	8,0450	2,3203	0,0665	0,5020
		SD	0,0	0,2609	1,0670	0,3973	0,0102	0,0075
		CV	0,0	0,1185	0,1326	0,1712	0,1537	0,0149
	93-88	Media	5,14	2,9698	6,0601	0,7264	0,1007	0,4602
		SD	0,2	0,4192	0,8681	0,3685	0,0192	0,0112
		CV	0,0	0,1412	0,1432	0,5073	0,1902	0,0244

Las estadísticas descriptivas para las variables químicas que se pueden observar en la Tabla 16, muestran que en la variable pH, el café tipo A (93-88 °C), el tipo B (90-90 °C) presentaron los promedios más altos con valores de pH 5.7 y el café B (93-88 °C) siguió con 5.6; por el contrario el café A (88-93 °C) y el Café C (93-88 °C) enseñaron los valores de pH más bajo con 5.0 y 5.1 respectivamente. Para la

variable índice de color, el café A (88-93 °C) obtuvo el índice de color más alto con 5.79, siguiendo el mismo café A (90-90 °C) con 5.102, en contra parte el café B (93-88 °C) y C (90-90 °C) obtuvieron los índices de color más pequeños con 2.00 y 2.201. En los polifenoles los cafés C (88-93 °C), A (88-93 °C) y B (88-93 °C) con 15.1529, 1.1633 y 11.7883 g/L respectivamente obtuvieron los valores más altos en comparación con los otros cafés a diferentes temperatura, en contraste, los valores más bajos fueron para los cafés C (93-88 °C) con 6.06 g/L, A (93-88 °C) con 7.3694 g/L y B (90-90 °C) con 7.3711 g/L.

Con relación a la cafeína se puede decir que los promedios más altos, los presentaron los cafés A (88-93 °C), B (90-90 °C), A (90-90 °C) y C (90-90 °C) con 3.8094 g/L, 2.5591 g/L, 2.4451 g/L y 2.3206 g/L respectivamente y los valores más bajos se obtuvieron en el café C (93-88 °C) con 0.7264 g/L y A (93-88 °C) con 1.1947 g/L. En cuanto al contenido de lípidos, el café B (90-90 °C) con 0.1084 g/L, C (93-88 °C) con 0.1007 g/L, y A (90-90 °C) con 0.0871 g/L mostraron los valores más altos, contrario al café A (93-88 °C) que presentó el valor más bajo con 0.0326 g/L. Finalmente para las variables químicas, el ácido clorogénico mostró que el café C (88-93° C) con 0.05553 g/L fue el contenido promedio más alto que el resto de cafés y en donde el valor más bajo fue de 0.3131 g/L para A (93-88 °C).

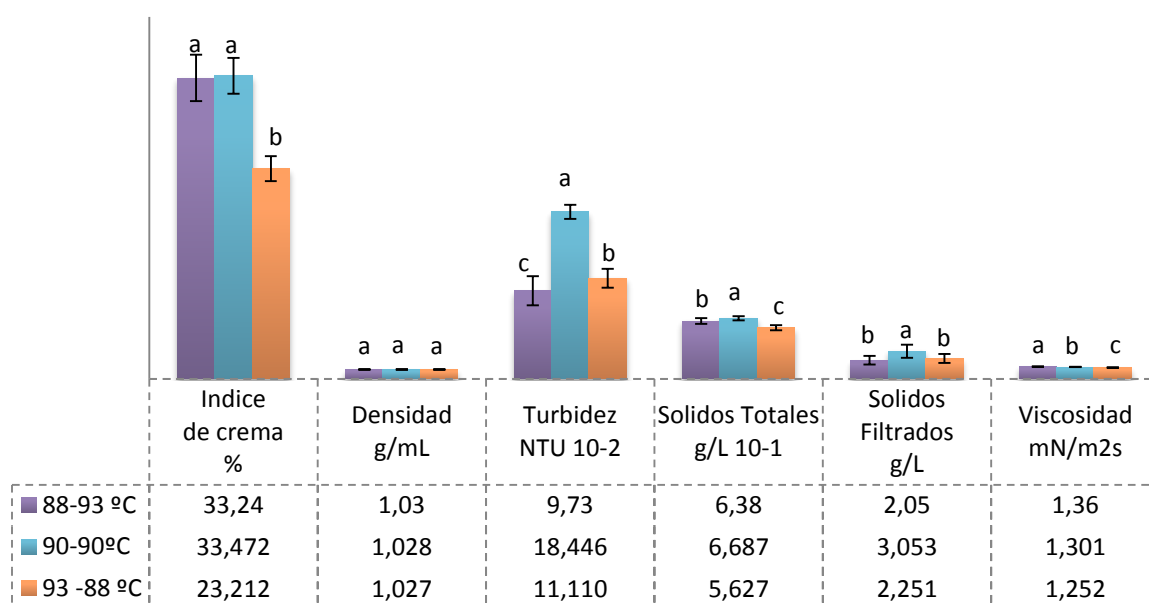
Los coeficientes de variación (CV) que evalúan la calidad estadística de las estimaciones, muestran que para la mayoría de las variables fueron menores del 20%, es decir, que su precisión fue admisible y por lo tanto las determinaciones realizadas fueron muy uniformes.

Sin embargo, para variables físicas el CV Los coeficientes de variación fueron superiores al 20%. En las **variables físicas** para sólidos filtrados en el café A (88-93 °C), A (90-90 °C), A (93-88 °C) y B (93-88 °C), así como para la turbidez en el café B (88-93 °C). Para **variables químicas** los polifenoles del café B (88-93 °C), la cafeína en los cafés A (93-88 °C), C (88-93 °C) y C (93-88 °C) y finalmente en la determinación de los lípidos del café A (93-88 °C). Estos valores indican que posiblemente hubo factores externos al proceso de medición que influyeron en el resultado final de las determinaciones dado que todas las muestras en todos los

café y todas las temperaturas fueron tratados de la misma manera, estos factores podrían ser ocasionados por la propia naturaleza del café, por los perfiles o curvas de temperatura empleadas y/o por la granulometría.

4.2.3 PROPIEDADES FÍSICAS POR CAFÉ OBTENIDOS A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN

En las Figuras 15 a 17 que se presentan a continuación se podrá ver la variación de las propiedades físicas como el índice de crema, la densidad, la turbidez, los sólidos totales, los sólidos filtrados y la viscosidad con relación a las diferentes curvas de temperatura de extracción ascendente, plana y descendente por cada tipo de café. Las letras iguales no presentan diferencias significativas en la prueba Tukey (10%).



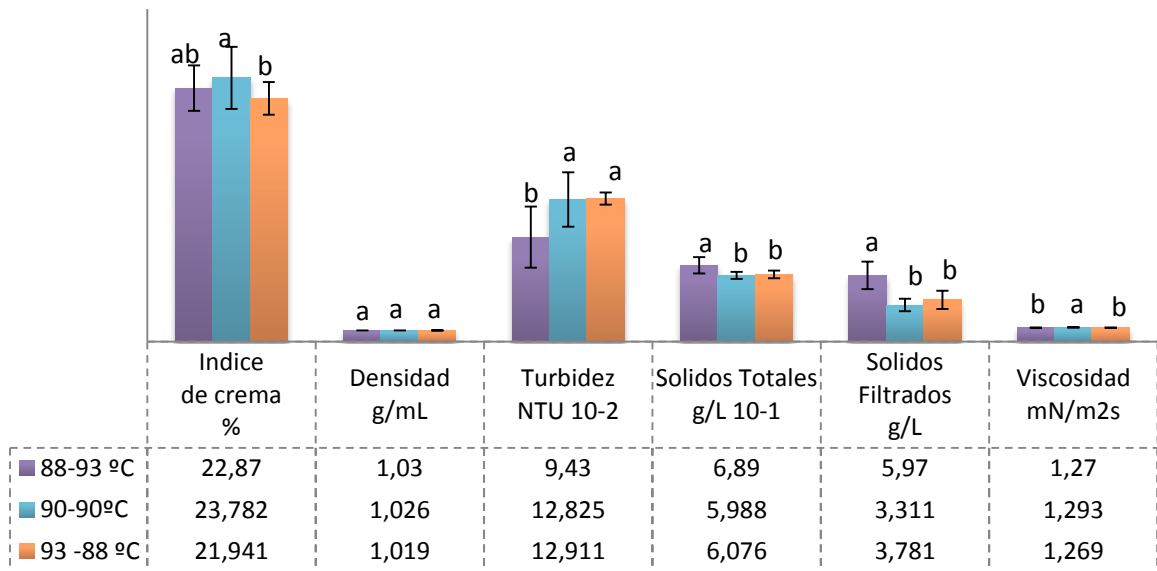
Valores medios \pm desviación estándar (n = 18). Prueba Tukey ($\alpha=0.1$): Letras distintas indican diferencias significativas.

Figura 15. Propiedades Físicas del Café tipo A (Robusta Natural).

En el café tipo A (Robusta Natural) representado en la Figura 15, se puede observar que el índice de crema, turbidez, sólidos totales y sólidos filtrados presentaron los valores más altos de estas variables cuando el perfil de temperatura fue plano (90-90

°C). En el perfil de temperatura ascendente (88-93 °C) la viscosidad fue más alta en comparación con los otros dos perfiles. En la curva de temperatura descendente (93-88 °C) se mostraron los valores más pequeños de las variables, índice de crema, sólidos totales y viscosidad.

Para el café tipo B (Arábica Natural) en la Figura 16, la curva de temperatura plana se observan los valores más altos de las variables Índice de crema, turbidez y viscosidad. Los sólidos totales y sólidos filtrados fueron mayores en la curva ascendente.



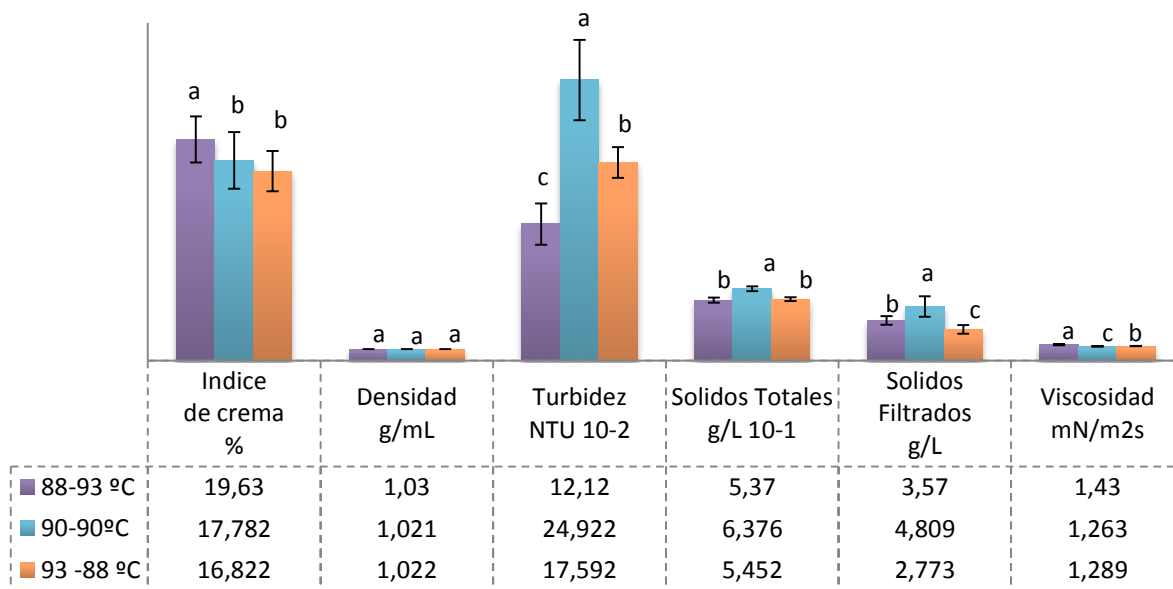
Valores medios ± desviación estándar (n = 18). Prueba Tukey ($\alpha=0.1$): Letras distintas indican diferencias significativas.

Figura 16. Propiedades Físicas del Café Tipo B (Arábica Natural).

En los perfiles de temperatura plana y ascendente se pudo observar que no hubo diferencias significativas entre las variables, turbidez, sólidos totales y sólidos filtrados, pero es de resaltar el patrón de comportamiento que se presentó en dichas variables.

El café tipo C (Arábica Lavado) de la Figura 17, mostró los mayores valores de las variables índice de crema y viscosidad cuando la curva de temperatura fue

ascendente. En la curva de temperatura plana, se observa que las variables turbidez, sólidos totales y sólidos filtrados son mayores que en el resto de los perfiles. Los valores intermedios en las variables se produjeron cuando la curva de temperatura fue descendente.



Valores medios \pm desviación estándar (n = 18). Prueba Tukey ($\alpha=0.1$): Letras distintas indican diferencias significativas.

Figura 17. Propiedades Físicas del Café Tipo C (Arábica Lavado).

4.2.4 COMPARACIÓN DE ALGUNAS PROPIEDADES FÍSICAS DEL CAFÉ A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN

ÍNDICE DE CREMA

Según Illy y Viani (2005), la crema del café espresso, es principalmente dióxido de carbono generado durante el tostado de café y atrapado dentro de la estructura celular, que cuantitativamente debe representar al menos el 10% del volumen de un espresso y tener una persistencia hasta los 40 minutos. En este caso, tanto el índice de crema (>16%) como la persistencia (>10 min) sobrepasaron estos valores en todos los cafés bajo diferentes curvas de temperatura, aunque técnicamente una

persistencia mayor a 8 min no es interesante para el consumidor, ya que es una bebida que se toma inmediatamente después de su elaboración.

Además, la crema es una capa espumosa que ejerce una barrera selectiva contra los volátiles. Por lo tanto el espesor y color de la crema podría modificar los compuestos volátiles de la taza que pueden implicar o no una aceptación sensorial (Parenti et al. 2014). Aunque, aquí no se puede describir por completo el aroma percibido, el color de la crema pudo determinar que los espressos no fueron sobre-extraídos y por lo que es posible que la abundancia de algunos compuestos, podrían incluso dar lugar a un olor o sabor agradable.

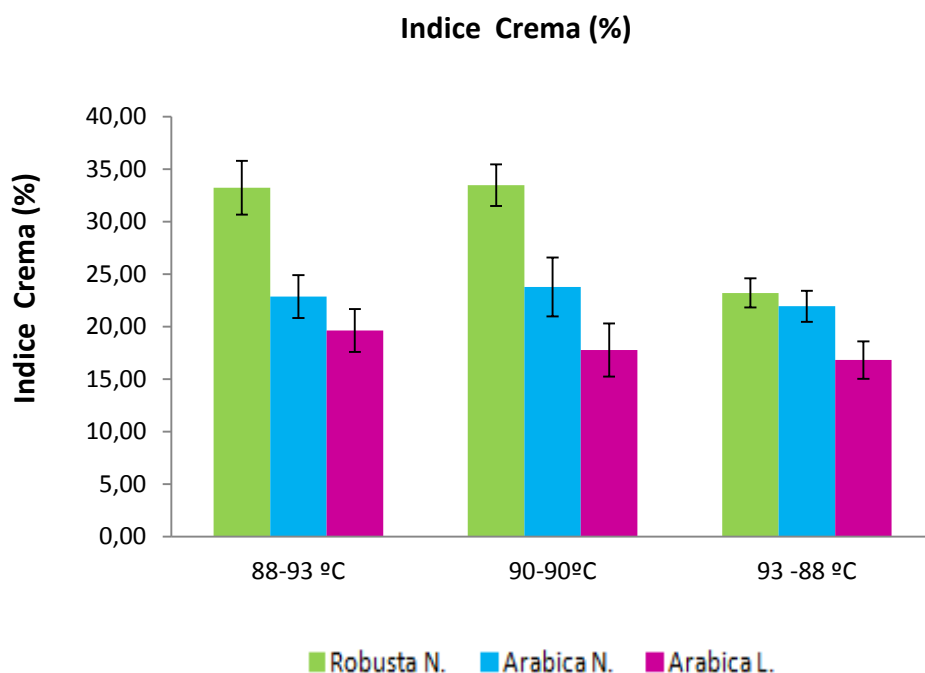


Figura 18. Índice de crema por café y por temperatura.

Los valores altos en el índice de crema de los tres cafés de la Figura 18, en general descansan en la hipótesis de que no hay pérdida de dióxido de carbono debida a la rápida preparación de café espresso inmediatamente después de la molienda. Según Illy & Navarini (2011) mencionan que el rango de 0-30 min es el intervalo entre la molienda y la preparación de café espresso, y no debe superarse con el fin de preservar la calidad. Por otro lado, el café espresso elaborado con café Robusta

proporcionó índices más altos de crema que los extractos de café Arábica, lo que se corrobora en el mismo documento por Bazzara (2004), indicando que la crema del café Robusta tiene mayor cantidad burbujas que la de café hecho con Arábica. El índice de crema de los cafés Arábica se ven menos afectados por los cambios de temperatura que el café Robusta, que mostró una disminución, cuando el perfil de temperatura fue descendente.

DENSIDAD

La densidad en la bebida del café es la relación que existe entre la masa y el volumen de un líquido. En la Figura 19, se muestran las densidades obtenidas para los tres tipos de café a diferentes perfiles de temperatura. Como se observa no hubo diferencias notables en ninguno de los cafés bajo las diferentes curvas de temperatura, situación que se corrobora con estudios realizados por Andueza et al. (2003a, 2003b, 2007).

Masella et al. (2015), publicaron que la densidad viene afectada significativamente por la presión y no por la temperatura, lo que demuestra que los perfiles de temperatura no modifican la densidad.

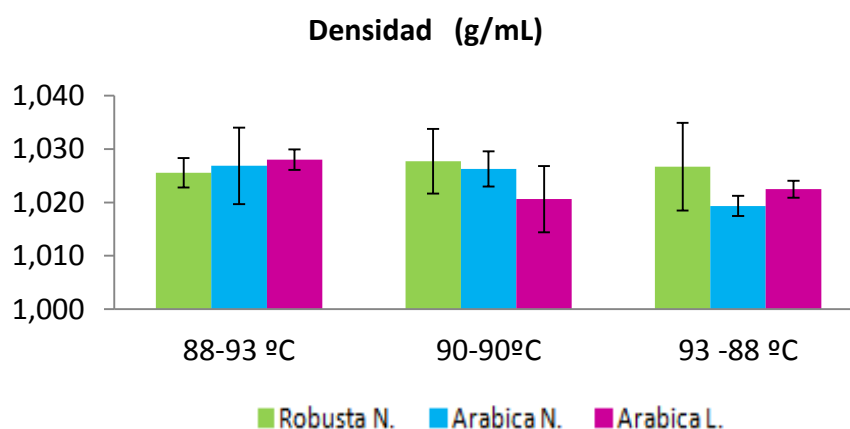


Figura 19. Densidad por café y por temperatura.

TURBIDEZ, SÓLIDOS TOTALES, SÓLIDOS FILTRADOS Y VISCOSIDAD

La turbidez, que se define como la falta de transparencia de un líquido debido a la presencia de partículas en suspensión, está relacionada directamente con los sólidos totales, sólidos filtrados y la viscosidad. Por eso en la Figura 20 se pueden observar estas cuatro variables por tipo de café y por perfil de temperatura.

Estas variables son las características más sobresalientes del espresso, a menudo percibida por los consumidores como "fuerte" y por los catadores como "cuerpo". Los sólidos totales y los sólidos filtrados, están determinados por el secado de una muestra del extracto en un horno, hasta peso constante con y sin filtrado. Según Illy & Viani (2005), los sólidos suspendidos, lípidos y solutos emulsionados son incluidos. La concentración de los sólidos totales y filtrados depende de la relación café/agua, y también depende del grado de tostado, molido y la temperatura de percolación.

La viscosidad del café es considerablemente mayor que la del agua pura (1.00×10^{-2}), es afectada por la presencia de fases dispersas y relacionada con la cantidad de lípidos en emulsión. Una alta viscosidad está asociado una mayor crema, cuerpo, sabor y olor. (Illy & Viani, 2005).

Por otro lado, Masella et al. (2015), encontró que la viscosidad y sólidos totales no fueron significativamente afectados por la temperatura empleando temperaturas constantes de 75 °C y 85 °C, sin embargo, en este estudio empleando los perfiles de temperatura de extracción, se observaron diferencias y esto es muy importante ya que los consumidores asocian generalmente este parámetro con la fuerza del café.

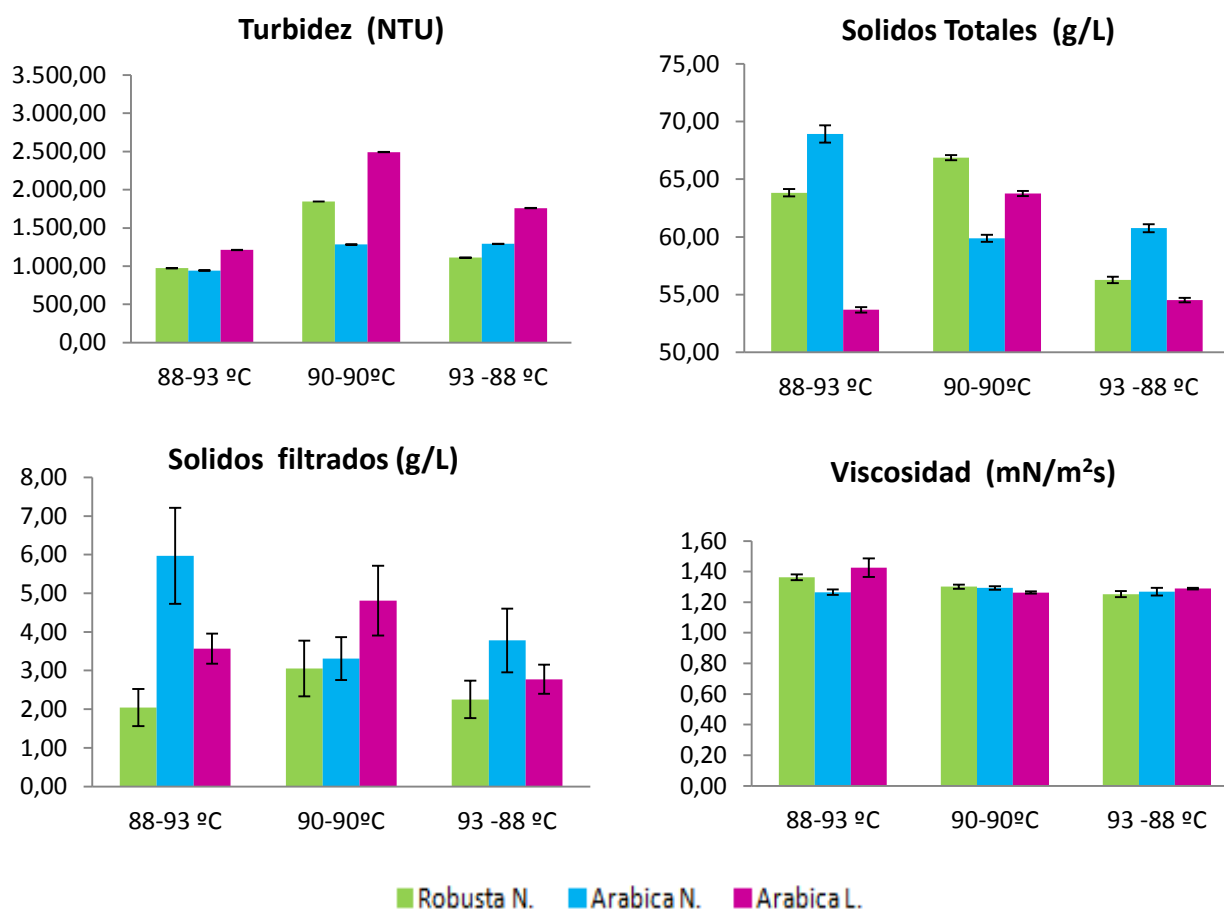


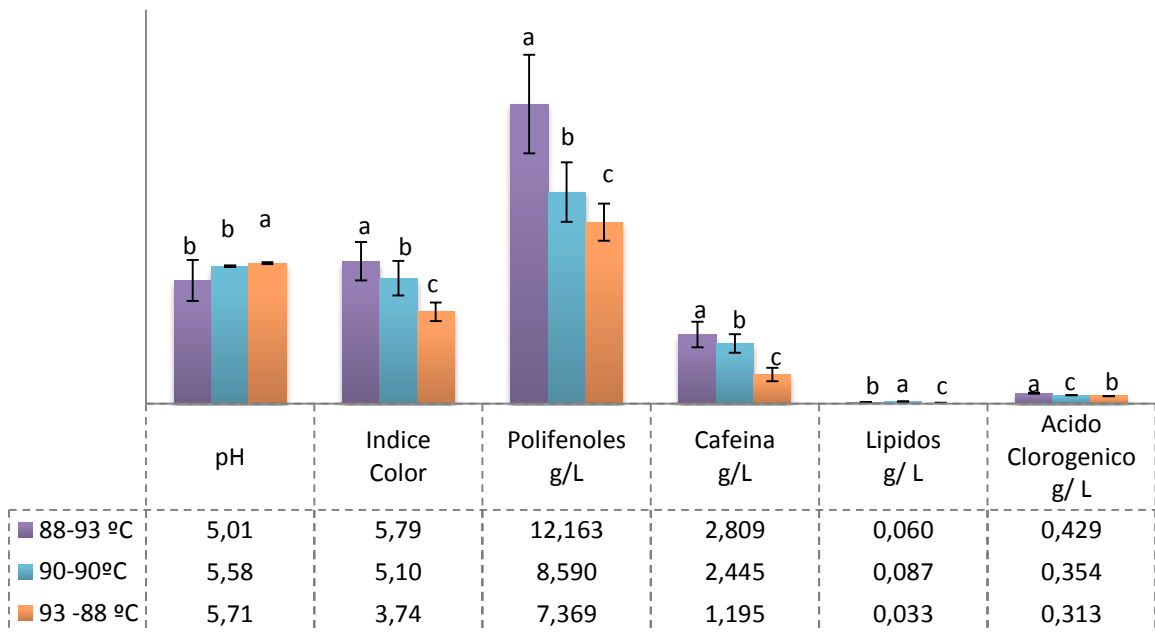
Figura 20. Turbidez, Sólidos Totales, Sólidos filtrados y Viscosidad por café y por temperatura.

4.2.5 PROPIEDADES QUÍMICAS POR CAFÉ A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN

En las siguientes tres Figuras (21- 23), se observará como varían las propiedades químicas como el pH, el índice color, los polifenoles, la cafeína, los lípidos y el ácido clorogénico cuando se emplean los tres perfiles de temperatura de extracción (88-93 °C , 90-90 °C y 93-88 °C).

Como se observa el café tipo A de la Figura 21, el índice color, los polifenoles, la cafeína y el ácido clorogénico tuvieron los valores altos cuando se utilizó la temperatura ascendente. Para la curva de temperatura descendente, las variables

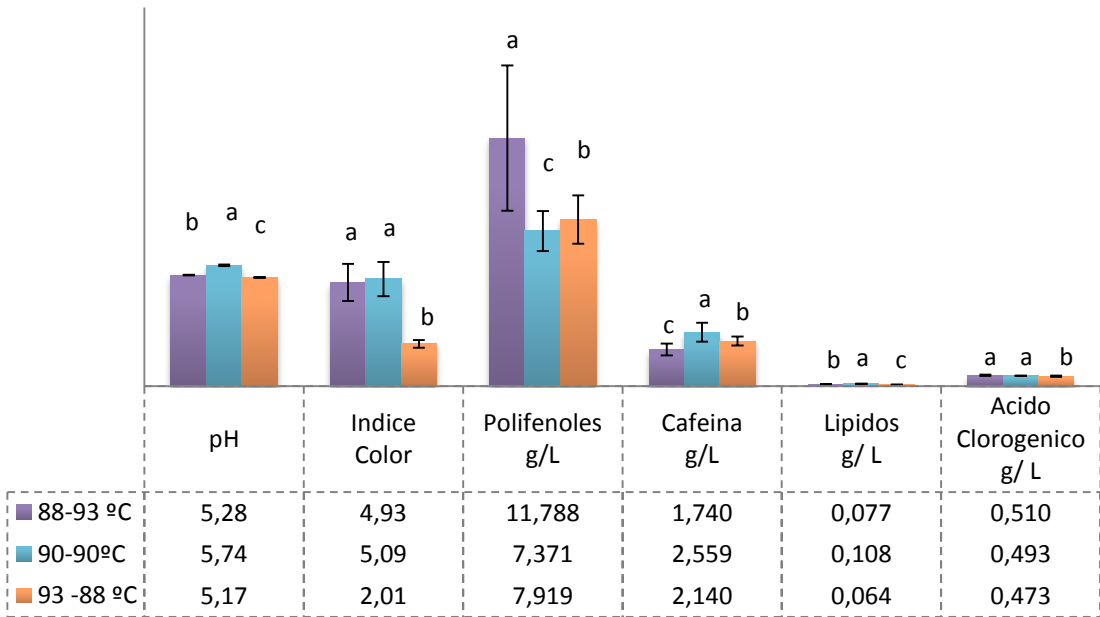
anteriores junto con los lípidos fueron más bajos en comparación con las otras curvas de temperaturas y su pH fue el más alto. En el perfil plano, se obtuvieron todos los valores intermedios excepto los lípidos donde su valor fue el más alto. Ahora, si se mira detalladamente la Figura, se puede ver un patrón de comportamiento de las variables en el café Tipo A cuando se varía la temperatura, es decir que las variables índice color, los polifenoles, cafeína y el ácido clorogénico tienden a disminuir cuando la temperatura aumenta.



Valores medios ± desviación estándar (n = 18). Prueba Tukey ($\alpha=0.1$): Letras distintas indican diferencias significativas.

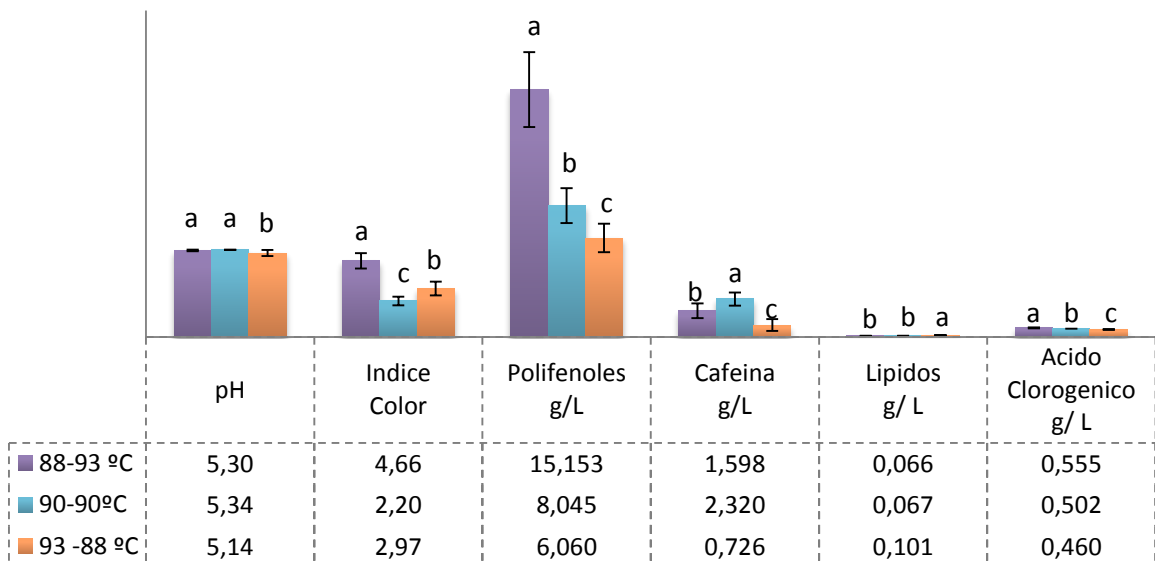
Figura 21. Propiedades químicas del Café Tipo A (Robusta natural).

Para el Café tipo B, en la Figura 22, los polifenoles se destacan del resto de las variables por presentar un valor muy alto en la temperatura ascendente con relación a las demás temperaturas. El pH, el índice color, cafeína, lípidos y ácido clorogénico tuvieron valores mayores en el perfil de temperatura plana. En el perfil descendente, el pH, el índice color y los lípidos fueron muy bajos.



Valores medios ± desviación estándar (n = 18). Prueba Tukey ($\alpha=0.1$): Letras distintas indican diferencias significativas.

Figura 22. Propiedades químicas del Café Tipo B (Arábica natural).



Valores medios ± desviación estándar (n = 18). Prueba Tukey ($\alpha=0.1$): Letras distintas indican diferencias significativas.

Figura 23. Propiedades químicas del Café Tipo C (Arábica Lavado).

Por otro lado, el café tipo C de la Figura 23, muestra que el pH, índice color, polifenoles y ácido clorogénico tuvieron mayores valores en el perfil de temperatura ascendente. Las variables pH, polifenoles, cafeína y ácido clorogénico mostraron los valores más bajos en el perfil descendente.

4.2.6 COMPARACIÓN DE ALGUNAS PROPIEDADES QUÍMICAS DEL CAFÉ A DIFERENTES CURVAS DE TEMPERATURAS DE EXTRACCIÓN

EL pH

En la Figura 24, muestra el pH de los tres tipos de café a diferentes perfiles de temperatura. El pH del café se relaciona con la percepción de la acidez según lo demuestra Boeneke et al. (2006), quien concluye que el tostado del café influye en los ácidos clorogénico y los sólidos orgánicos que se transforman en otros productos a causa del calor, disminuyendo el contenido de ácido y aumentando el pH.

Para Masella et al. (2015) los valores de pH más altos están relacionados probablemente cuando los cafés espresso se producen a altas presiones y no a la temperatura. Sin embargo, en este caso se observa que la temperatura tiene un efecto en el pH. En los cafés Arábicas se observa una tendencia en los valores del pH, en donde son altos en el perfil de temperatura plana y menor en los perfiles de temperatura ascendente y descendente siendo este último el de menor pH. Por otro lado, el café Robusta presenta una sensibilidad marcada con los cambios de temperatura, en donde a una curva de temperatura descendente el pH aumenta y cuando la curva es descendente este tiende a disminuir el pH.

Humphries (2008) explica que existen muchos tipos de acidez presentes en una taza de café, cada uno con su propia composición molecular y cada uno con diferentes sabores. La acidez del café puede expresarse en dos términos, la “acidez orgánica” compuesta por algunos ácidos cítricos, ácidos málicos, ácidos lácticos y ácidos acéticos que al combinarse con los azúcares innatos y carbohidratos en el café tostado proveen un sabor placentero a la taza de café y por la “acidez clorogénica” compuesta por algunos ácidos como el caféico y el ácido quínico que son dos

grandes tipos de ácidos clorogénicos, presentes en cualquier taza de café en proporciones variables. Ambos tienen sabores desagradables: amargos, intensamente duros, y comúnmente metálicos. Esta última categoría es la que usualmente lleva a las personas a relacionar el café con el término “acidez”, es decir que cuando hay mayor presencia de ácidos clorogénicos, el pH de café tiende a disminuir.

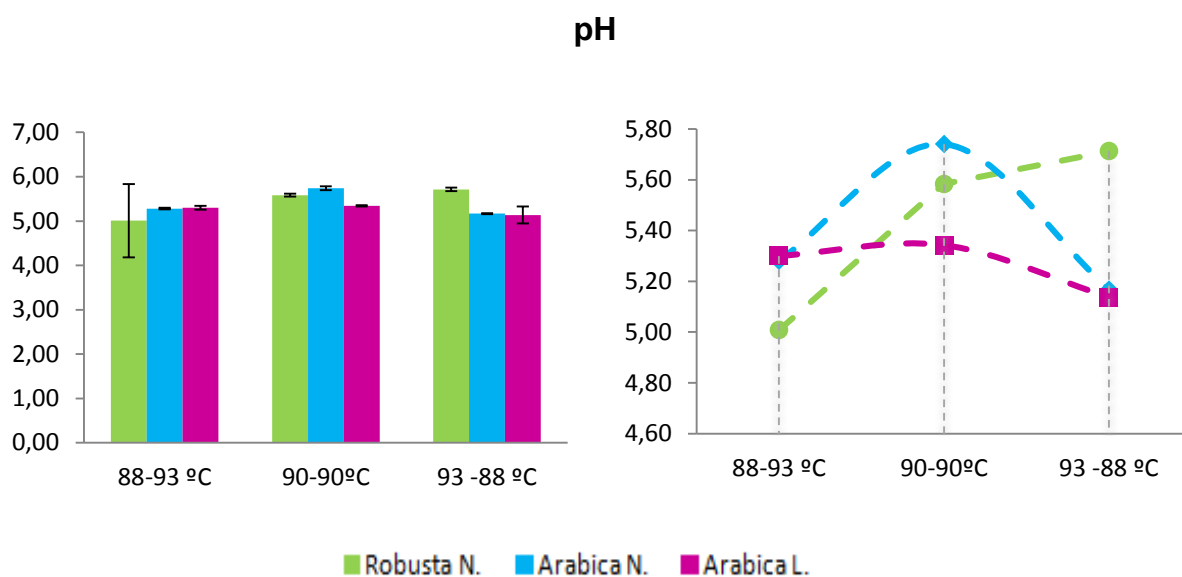


Figura 24. pH por tipo de café y por temperatura.

POLIFENOLES, ACIDO CLOROGÉNICO Y CAFEÍNA

Los polifenoles son responsables del color, la astringencia y del gusto amargo del café; se encuentran en varias estructuras químicas y juegan un papel importante en la calidad y en la salud por su fuerte acción antioxidante (Charurin et al. 2002).

Los polifenoles se dividen en flavonoides (ácidos fenólicos y estilbenos) y no flavonoides (antocianas, flavonoles y taninos). Dentro de los ácidos fenólicos se encuentra el ácido caféico, su activo principal es el café y está presente en los procesos de oxidación de compuestos como en el caso del vino. Los ácidos cumáricos y felúricos, son importantes a nivel organoléptico ya que se les atribuye

los sabores amargos. Las antocianinas, flavonoles y taninos han demostrado poseer una elevada capacidad antioxidante en diversas investigaciones en café, aunque a los taninos también se les atribuye el característico sabor astringente que puede ser enmascarado al formar soluciones estables con los polisacáridos (Marín et al. 2008; Catania & Avagnina, 2007; Nicoli et al. 1997; Puertas et al. 2012; Hagerman et al. (1998).

Como se observa en la Figura 25, los polifenoles tuvieron un patrón de variación similar en los tres tipos de café, lo que muestra su sensibilidad a los cambios de temperatura. El mayor contenido de polifenoles se presentó en el café Arábica Lavado para el perfil de temperatura ascendente, en el Robusta Natural para la curva plana y en el Arábica Lavado para el perfil de temperatura descendente. Según Capel et al. (2010) la eficacia antioxidante del café Robusta es superior al del Arábica, lo que se corrobora con el alto contenido de polifenoles obtenido en el perfil de temperatura plano.

Los ácidos clorogénicos del café involucran alrededor de 40 ácidos, Illy & Viani (2005), señalan que la mayoría de los ácidos orgánicos que están presentes en el café verde se comportan como los ácidos clorogénicos que se descomponen parcialmente en los ácidos cítricos y ácido málico para formar fumárico y maleico. Después de ser tostado el café estos ácidos se transforman probablemente en ésteres como el éster de ácido fumárico/ácido málico y otro monoéster como el del ácido málico. Un gran número de ácidos se genera por reacción de Maillard o caramelización (Castillo et al. 2002).

El ácido clorogénico más abundante en los géneros del café, es el clorogénico 5-cafeóil-quínico (5-CQA), como ha sido reportado por Marin & Puerta (2008) donde se estimó que este ácido constituyó del 87% al 94.5% del contenido de los ácidos clorogénicos presentes en los cafés Arábicas y el 87% del contenido de los ácidos clorogénicos presentes en los cafés Robustas.

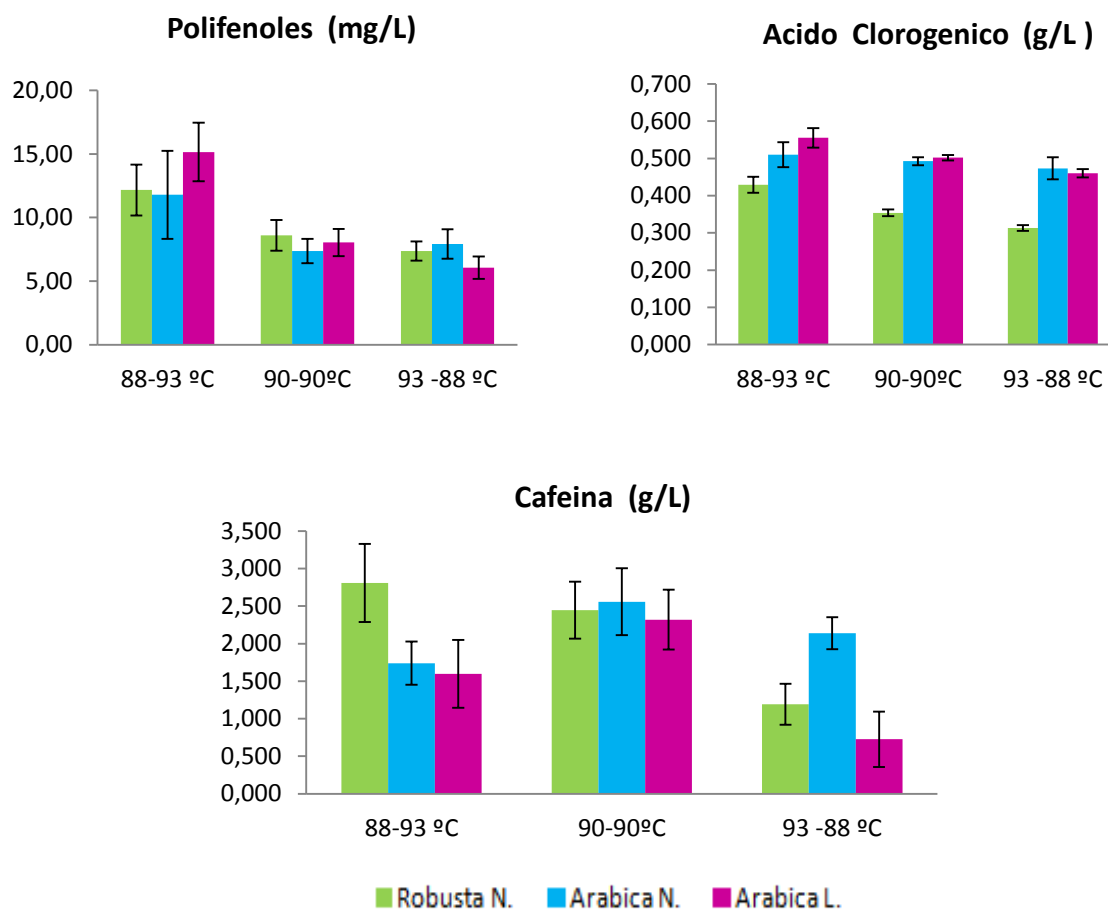


Figura 25. Polifenoles, Acido Clorogénico y Cafeína por tipo de café y por temperatura.

Según Puertas (2011b) los granos de Robusta se destacan por presentar el mayor contenido de polisacáridos, cafeína, ácidos clorogénicos y cenizas que el café Arábica. Por otra parte Richelle et al. (2001) afirmación que el café Robusta duplica al café Arábica en su capacidad antioxidante, por su mayor contenido en ácido clorogénico. Sin embargo y como se observa en la Figura 25, para todas las curvas de temperatura, los espressos elaborados con café Robusta Natural presentaron los valores más bajos en contraste con los dos cafés Arábicas (Natural y lavado), demostrando así, que la temperatura de extracción influye en el contenido final de los ácidos clorogénicos.

La cafeína es un tipo de alcaloide que contribuye al sabor amargo del café según

Fox, et al. (2013). El contenido de cafeína extraída en taza, depende del tipo de café y del método de preparación. El café Robusta presenta mayor contenido de cafeína que un Arábica. El café Arábica beneficiado por vida seca (Arábica natural) presenta mayor contenido de cafeína extraída que el beneficiado por vía húmeda (Arábica lavado). Por otro lado, en el café espresso empleando máquinas de café convencionales, la extracción de cafeína está al alrededor del 40% en los primeros 5 segundos de erogación y se ha calculado que en los primeros 10 segundos más del 60% de la cafeína es extraída (Puertas 2011a; Mora & Rodríguez, 2010).

En la Figura 25, se observa exactamente lo mencionado previamente, cuando la temperatura es uniforme, es decir se emplea un perfil de temperatura plana, el contenido de cafeína es mayor en todos los tipos de café, especialmente en el Arábica Natural y en el Robusta Natural. Cuando hay modificación en la curva de temperatura, o sea, cuando se emplean perfiles de temperatura ascendente y descendente, se muestra que en los cafés Robusta Natural y Arábica Natural, se obtuvo un mayor contenido de cafeína en estas curvas respectivamente. El café Arábica Lavado presentó siempre los menores niveles de cafeína en todos los perfiles de temperatura.

Finalmente se puede decir que el contenido de cafeína, ácido clorogénico y polifenoles varía según los perfiles de temperatura independiente del tipo de café, aunque Masella et al. (2014) publicó que la cafeína, y ácidos clorogénicos no fueron afectados por los valores de presión o de temperatura en los intervalos 75 – 85 °C.

LÍPIDOS

La mayoría de los lípidos del café se encuentran en el endospermo y sólo una pequeña cantidad denominada cera está en la capa externa del grano. En la taza, la cantidad de lípidos depende de varios aspectos como el tipo de café, tostado, molido, calidad del agua, presión, temperatura y tiempo.

El café espresso proporciona una buena extracción de lípidos en la bebida. Estos son excelentes portadores de aroma, por lo que este proceso sería una manera

óptima de incrementar el aroma en taza (Illy & Navarini, 2005; Speer & Kölling, 2006). Además, la concentración de lípidos afecta fuertemente a la fase de la crema del espresso (Parenti et al. 2014).

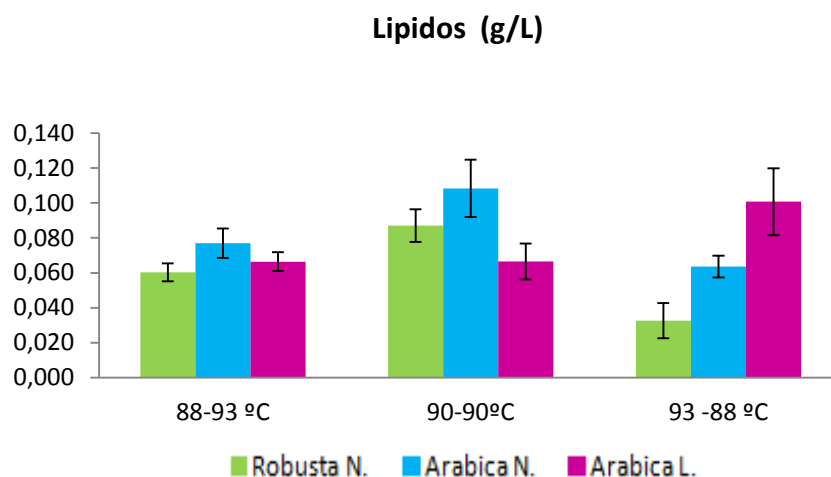


Figura 26. Lípidos por tipo de café y por temperatura.

En la Figura 26, se puede ver que hay un patrón de comportamiento en los cafés que han sido beneficiados por vía natural. Esto puede deberse tanto a la composición química del mucilago como a su acción como película protectora frente al tueste del café. Además, se observa que en los perfiles temperatura ascendente y plana el café Arábica Natural tiene el mayor contenido de lípidos totales, mientras que en el perfil de temperatura descendente el mayor valor fue para el Arábica Lavado, comportamiento que es de esperarse en el Arábica donde el contenido de lípidos en el espresso se encuentran presentes en mayor cantidad que en el café Robusta (Puertas 2011b).

Finalmente, los polifenoles, el ácido clorogénico y la cafeína en el café espresso son independientes de los contenidos iniciales en grano y su cantidad pueden ser modificados por el perfil de temperatura en el proceso de extracción. El análisis de estas variables puede ser indicador del gusto del café como fue dicho previamente. La Organización Internacional del Café - OIC (2010) mencionó que el café Arábica desde el punto de vista organoléptico es más ácido y menos amargo que el Robusta,

esto puede deberse a que la cantidad de polifenoles puede enmascarar sustancias permitiendo expresar los ácidos clorogénicos, mientras que la temperatura ejerce una acción contundente en la cafeína, responsable en gran medida del gusto amargo del café.

Los lípidos, según Puertas (2011b) ayudan al transporte de los aromas y sabores y en el espresso dan sabor y cuerpo. Para Illy & Navarini (2011), el contenido de lípidos está relacionado con la estabilidad de la crema, por ejemplo, el espresso de café Arábica presenta un mayor contenido de lípidos poco estables lo que causa ruptura del equilibrio en la crema, sin embargo, el espresso elaborado con Robusta presenta lípidos emulsionados muy estables lo que hace presenciar una crema más compacta.

4.2.7. RELACIÓN DE LAS PROPIEDADES FISICOQUÍMICAS CON LAS PROPIEDADES ORGANOLÉPTICAS

Con los análisis bibliográficos realizados previamente en la sección 4.2, se pudo realizar una relación entre las propiedades fisicoquímicas y las propiedades organolépticas de los tres cafés a diferentes perfiles o curvas de temperaturas de extracción.

Inicialmente, se pudo inferir que el café espresso producido con Robusta Natural, empleando la curva de temperatura ascendente, permite describir un espresso de cuerpo ligero y sabor fuerte, mientras que con la curva descendente se puede obtener un espresso con un buen índice de crema, un cuerpo ligero, con menor acidez y de sabor agradable.

El espresso obtenido con Arábica Natural, reveló que en el perfil de temperatura ascendente permite obtener un espresso de mayor cuerpo, con sabor más intenso y balanceado si se compara con el perfil de extracción plana. Por otro lado, empleando el perfil de temperatura descendente se deriva un espresso de buen cuerpo y sabor más ligero comparado con los espressos obtenidos mediante las curvas ascendente y plana.

Finalmente, los espressos obtenidos a partir de Arábica Lavado empleando el perfil de temperatura ascendente dieron lugar a un café espresso de cuerpo ligero y sabor intenso, con un aparente balance entre ácidos-amargos o con una ligera tendencia hacia amargos, si los comparamos con los que se obtuvieron con el perfil de temperatura plana. Con la curva de temperatura descendente, los espressos obtenidos se pueden describir como menos intensos y con una tenue acidez y un cuerpo notable.

4.3. ELEMENTOS DISCRIMINATORIOS PARA CARACTERIZAR EL CAFÉ ESPRESSO A TRAVÉS DE DESCRIPTORES ORGANOLÉPTICOS Y PROPIEDADES FÍSICOQUÍMICAS

Los resultados que se muestran a continuación, son el producto del análisis realizado para las variables fisicoquímicas y para los descriptores organolépticos antes de realizar ACP para los descriptores organolépticos.

4.3.1 ESTADÍSTICAS DESCRIPTIVAS DE LAS VARIABLES

En la Tabla 17, se pueden observar las estadísticas descriptivas básicas (media, desviación estándar y coeficiente de variación) de cada uno de los tres tipos de café (A = Robusta Natura, B= Arábica Natural y C= Arábica Lavado) en cada perfil o curva de temperatura de extracción (ascendente= 88-93 °C, plana= 90-90 °C y descendente= 93-88 °C).

Las **variables fisicoquímicas** como pH, color, polifenoles cafeína, lípidos ácido clorogénico, índice crema, turbidez, sólidos totales, sólidos filtrados y viscosidad, así como también los **descriptores organolépticos** que comprenden las variables intensidad del color de crema, textura de la crema, intensidad de olor, intensidad de tostado, cuerpo, acidez, amargo y astringencia, pueden ayudar a explicar un poco la calidad, además de permitir relacionar los resultados obtenidos.

Si se observa en detalle la Tabla 17, se muestra que tanto los promedios como las desviaciones estándar respectivas de cada una de las variables, permiten plantear que evidentemente hay una modificación en las propiedades fisicoquímicas y organolépticas en los diferentes espressos cuando se modifica la temperatura de extracción.

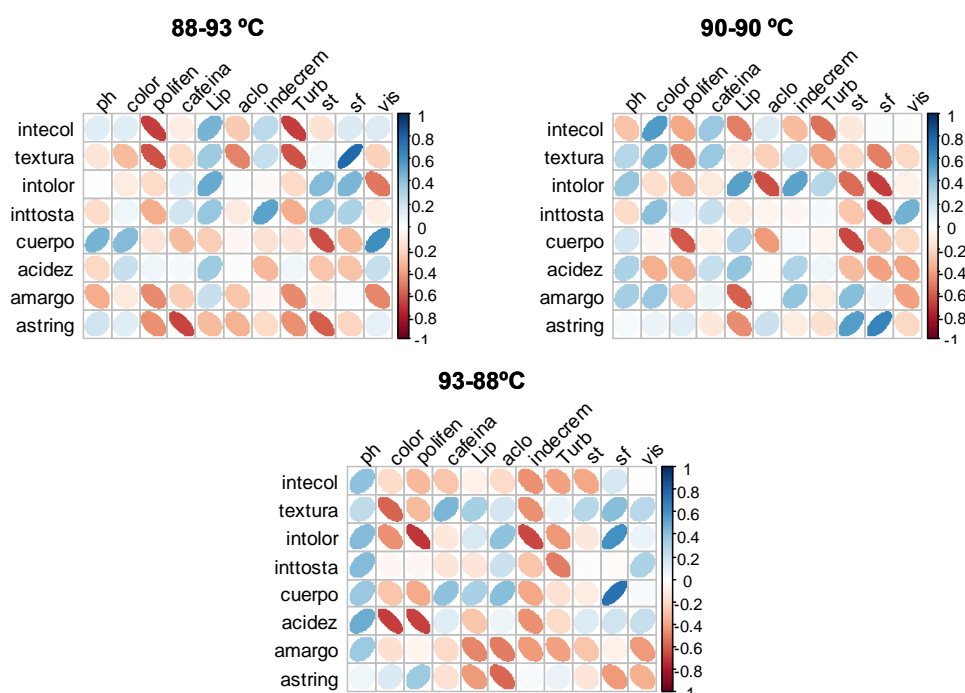
Ahora si se observa en particular el coeficiente de variación (CV) de todas las variables, se puede ver que es muy fluctuante de café a café y de perfil a perfil de temperatura. Por ejemplo, si se toma como referencia el café Arábica Natural y se selecciona variable color, cuando se emplea un perfil de temperatura ascendente el CV es del 18%, cuando se utiliza el perfil plano es de 13% y cuando se usa el perfil descendente es del 8%. Esto ocurre básicamente en el resto de variables que se presentan en la Tabla.

Tabla 17. Media, desviación estándar y Coeficiente de variación de los parámetros físicos químicos para los perfiles o curvas de temperatura.

PARAMETROS	Ascendente 88-93 °C									Plana 90-90 °C									Descendente 93-88 °C								
	Arábica Natural			Arábica Lavado			Robusta Natural			Arábica Natural			Arábica Lavado			Robusta Natural			Arábica Natural			Arábica Lavado			Robusta Natural		
	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV	Media	D. E.	CV
pH	5,29	0,02	0,00	5,26	0,03	0,01	5,19	1,01	0,19	5,72	0,05	0,01	5,36	0,01	0,00	5,57	0,02	0,00	5,17	0,01	0,00	5,24	0,33	0,06	5,72	0,05	0,01
Índice Color	4,858	0,861	0,18	4,24167	0,201	0,05	5,531	0,651	0,12	4,795	0,624	0,13	2,49	0,065	0,03	4,992	0,7468	0,15	2,054	0,1597	0,08	2,855	0,431	0,15	3,618	0,407	0,11
Polifenoles g/L	11,9532	3,422	0,29	17,4897	1,472	0,08	12,8533	1,58	0,12	7,0258	1,016	0,14	8,45967	1,032	0,12	8,8701	1,2286	0,14	7,8025	1,1541	0,15	6,16117	0,966	0,16	7,1679	0,66	0,09
Cafeína g/L	1,6668	0,284	0,17	1,47417	0,303	0,21	2,5994	0,606	0,23	2,586	0,566	0,22	1,91133	0,172	0,09	2,1908	0,2153	0,10	2,2222	0,2124	0,10	1,16983	0,05	0,04	1,2239	0,33	0,27
Lípidos g/ L	0,0781	0,0047	0,06	0,0695	0,0037	0,05	0,0607	0,0065	0,11	0,1072	0,0188	0,18	0,0732	0,0098	0,13	0,0855	0,0100	0,12	0,0607	0,0051	0,08	0,0989	0,0208	0,21	0,0294	0,0093	0,32
Acido Clorogénico g/ L	0,4986	0,0345	0,07	0,5598	0,0411	0,07	0,4374	0,0227	0,05	0,4878	0,0082	0,02	0,5020	0,0069	0,01	0,3545	0,0113	0,03	0,4634	0,0333	0,07	0,4620	0,0093	0,02	0,3106	0,0077	0,02
Índice de la crema	23,61	1,65	0,07	18,63	1,73	0,09	32,31	2,99	0,09	22,12	2,60	0,12	18,01	1,23	0,07	32,28	1,39	0,04	21,83	1,56	0,07	18,28	0,88	0,05	22,84	1,47	0,06
Turbidez NTU	956,25	273,75	0,29	1399,17	117,79	0,08	1028,25	126,41	0,12	1224,25	250,21	0,20	2075,00	95,52	0,05	1866,25	96,63	0,05	1315,95	51,47	0,04	1745,42	133,35	0,08	1052,25	38,36	0,04
Sólidos Totales g/L	67,239	7,26	0,11	54,2633	2,668	0,05	64,25	2,459	0,04	59,49	3,534	0,06	64,495	3,67	0,06	67,13	0,9844	0,01	60,72	3,5686	0,06	54,9133	0,881	0,02	56,61	3,108	0,05
Sólidos Filtrados g/L	5,866	1,45	0,25	3,685	0,421	0,11	2,204	0,417	0,19	3,672	0,321	0,09	4,28833	0,657	0,15	3,275	0,7401	0,23	3,424	0,9562	0,28	2,65667	0,283	0,11	2,277	0,202	0,09
Viscosidad mN/m2s1	1,2741	0,013	0,01	1,5055	0,016	0,01	1,3718	0,018	0,01	1,2935	0,012	0,01	1,2645	0,009	0,01	1,3024	0,0164	0,01	1,2693	0,0317	0,02	1,288	0,005	0,00	1,258	0,016	0,01
Intensidad Color	6,60	0,97	0,15	6,50	0,84	0,13	6,50	1,18	0,18	6,90	1,37	0,20	6,50	0,84	0,13	6,40	0,70	0,11	7,10	0,88	0,12	6,67	0,82	0,12	6,60	0,52	0,08
Textura	7,30	0,95	0,13	7,17	1,72	0,24	6,50	1,27	0,20	7,30	1,34	0,18	6,50	1,76	0,27	6,60	1,35	0,20	7,20	1,14	0,16	6,83	2,04	0,30	6,70	1,25	0,19
Intensidad Olor	7,30	0,67	0,09	7,17	1,47	0,21	7,10	1,20	0,17	7,00	1,41	0,20	6,67	1,37	0,20	6,40	0,84	0,13	7,20	1,23	0,17	7,33	0,82	0,11	6,70	1,83	0,27
Intensidad Tostado	6,40	1,07	0,17	6,17	1,17	0,19	6,50	0,97	0,15	6,10	1,37	0,22	6,50	0,55	0,08	6,20	1,03	0,17	6,90	0,74	0,11	6,50	1,38	0,21	6,40	1,26	0,20
Cuerpo	7,50	0,71	0,09	7,17	0,75	0,11	7,70	1,25	0,16	6,70	1,25	0,19	6,83	0,98	0,14	7,00	1,05	0,15	6,60	1,17	0,18	6,67	1,37	0,20	6,90	1,10	0,16
Ácidez	5,80	0,63	0,11	6,50	0,55	0,08	3,60	0,84	0,23	5,10	0,88	0,17	7,17	0,75	0,11	4,60	1,17	0,26	6,30	0,95	0,15	6,83	0,75	0,11	5,50	0,71	0,13
Amargo	4,40	1,07	0,24	4,17	1,17	0,28	5,80	2,10	0,36	5,10	1,85	0,36	4,67	2,34	0,50	5,90	2,02	0,34	5,50	1,84	0,33	4,83	2,32	0,48	6,00	1,70	0,28
Astringencia	2,60	1,26	0,49	2,17	1,60	0,74	3,40	2,07	0,61	3,60	2,01	0,56	3,17	2,04	0,64	3,70	1,57	0,42	3,20	1,75	0,55	3,33	2,16	0,65	3,30	1,25	0,38
Olor +	2,60	1,26	0,49	2,17	1,60	0,74	3,40	2,07	0,61	3,60	2,01	0,56	3,17	2,04	0,64	3,70	1,57	0,42	5,50	2,17	0,40	6,83	1,33	0,19	5,60	2,27	0,41
Olor -	2,60	1,26	0,49	2,17	1,60	0,74	3,40	2,07	0,61	3,60	2,01	0,56	3,17	2,04	0,64	3,70	1,57	0,42	1,60	0,84	0,53	1,50	0,84	0,56	2,00	0,94	0,47

4.3.2 RELACIÓN ENTRE LOS DESCRIPTORES ORGANOLÉPTICOS Y LAS PROPIEDADES FÍSICOQUÍMICAS

En las Figuras 27 - 29 se muestra la correlación por tipo de café y perfil de temperatura, en el eje **X** las propiedades fisicoquímicas y en el eje **Y** están los descriptores organolépticos. Los colores de las elipses indican la correlación, si es positiva es de color azul y si es negativa de color rojo. La intensidad del color indica la intensidad de la correlación. Las abreviaciones del eje **X** corresponden a pH (pH), color (color), polifenoles (polifen), cafeína (cafeina), lípidos (lip), ácido clorogénico (aclo), índice crema (indecrem), turbidez (turb), sólidos totales (st), sólidos filtrados (sf) y viscosidad (vis). Por otra parte, las abreviaciones del eje **Y** corresponden a intensidad color (intecol), textura (textura), intensidad olor (intolor), intensidad tostado (inttostado), cuerpo (cuerpo), acidez (acidez) amargo (amargo) y astringencia (astringencia).

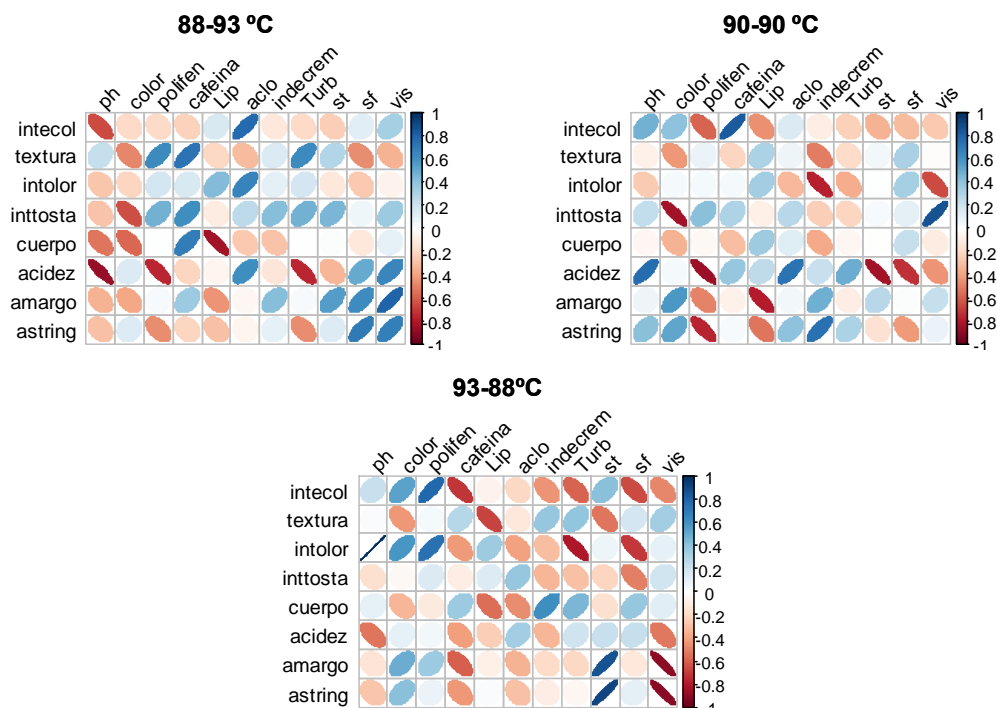


En el eje **X** las propiedades fisicoquímicas; en el eje **Y** están los descriptores organolépticos. Los colores de las elipses indican la correlación, si es positiva es de color azul y si es negativa de color rojo. La intensidad del color indica la intensidad de la correlación.

Figura 27. Correlaciones entre las propiedades fisicoquímicas y los descriptores organolépticos en el café Robusta Natural.

En la Figura 27 para el café Robusta Natural, la intensidad del color de la crema presenta una correlación negativa entre los polifenoles y la turbidez, en el perfil de temperatura ascendente y positiva con el índice color en la temperatura plana. La textura de la crema muestra una relación negativa con los polifenoles y la turbidez en el perfil ascendente, sin embargo la relación es positiva con los sólidos filtrados en el perfil de temperatura ascendente. La intensidad del olor presenta una correlación negativa con el ácido clorogénico, los sólidos filtrados y totales en el perfil plano. Sin embargo, en el perfil ascendente indica una correlación negativa con otras variables como los polifenoles y el índice de la crema. La intensidad del tostado en el perfil ascendente obtuvo una correlación positiva con el índice de la crema y negativa con los sólidos filtrados en el perfil de temperatura plana. El cuerpo, mostró una correlación negativa con los sólidos totales y positiva en el perfil ascendente; en el perfil plano mostró una correlación negativa entre los sólidos totales y los polifenoles y en el perfil descendente una relación positiva con los sólidos filtrados. La acidez, tuvo una relación inversa con el índice de color y los polifenoles en el perfil descendente. El amargor mostró una correlación negativa con la cafeína en el perfil de temperatura plana. Finalmente, la astringencia se destacó por tener una relación negativa con la cafeína en el perfil ascendente y positiva en el perfil plano con los sólidos totales y los sólidos filtrados.

En la Figura 28, para el café Arábica Natural, se observó que la intensidad del color de la crema en el perfil de temperatura ascendente tuvo una relación negativa con el pH y positiva con el ácido clorogénico, en el perfil plano mostró una correlación positiva con la cafeína y en la curva descendente, una asociación positiva con los polifenoles y negativa con la cafeína. La textura de la crema presentó una relación positiva con el contenido de polifenoles, cafeína y la turbidez, mientras que en el perfil descendente fue negativa con los lípidos. La intensidad del olor tuvo una correlación positiva con el ácido clorogénico en el perfil ascendente; en el perfil plano fue negativa con el índice de la crema y la viscosidad y en el perfil ascendente mostró una asociación positiva con el pH y negativa con la turbidez.

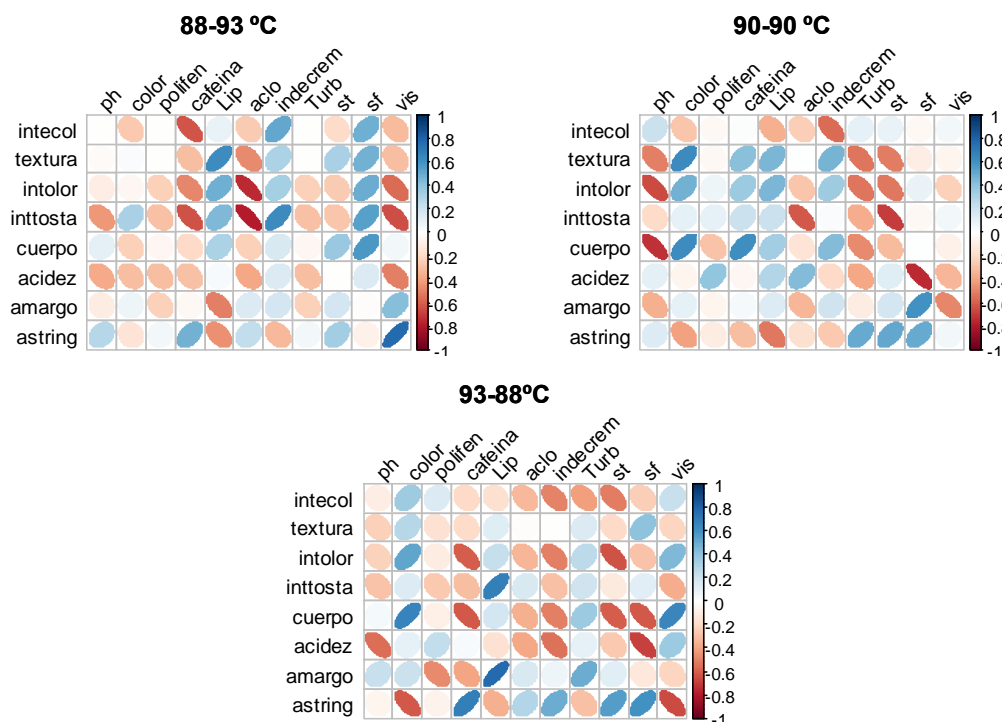


En el eje X las propiedades fisicoquímicas; en el eje Y están los descriptores organolépticos. Los colores de las elipses indican la correlación, si es positiva es de color azul y si es negativa de color rojo. La intensidad del color indica la intensidad de la correlación.

Figura 28. Correlaciones entre los descriptores organolépticos y las propiedades fisicoquímicas en el café Arábica Natural.

La intensidad del tostado se destacó por presentar una relación negativa con el índice color y positiva con la cafeína en el perfil ascendente y en el perfil plano tuvo una correlación negativa con el color y positiva con la viscosidad. El cuerpo mostró una relación positiva con la cafeína y negativa con los lípidos en la curva ascendente. Por otro lado, la acidez se destacó por mostrar una relación negativa con las variables pH, polifenoles y turbidez en la curva ascendente; en el perfil plano presentó una relación positiva con el pH y el ácido clorogénico mientras que fue negativa con los polifenoles, los sólidos totales y los sólidos filtrados. El amargor expuso una correlación positiva con la viscosidad en el perfil descendente; en el perfil de temperatura plana tuvo una relación negativa con la cafeína y en perfil descendente mostró una asociación negativa con la viscosidad y positiva con los sólidos totales. Finalmente, en el perfil ascendente la astringencia mostró una relación positiva con la viscosidad y los sólidos filtrados, en el perfil plano se destacó

por presentar una relación negativa con los polifenoles e inversa con en el índice de la crema y en el perfil descendente tuvo una relación positiva con los sólidos totales y negativa con la viscosidad.



En el eje X las propiedades fisicoquímicas; en el eje Y están los descriptores organolépticos. Los colores de las elipses indican la correlación, si es positiva es de color azul y si es negativa de color rojo. La intensidad del color indica la intensidad de la correlación.

Figura 29. Correlaciones entre los descriptores organolépticos y las propiedades fisicoquímicas en café Arábica Lavado.

En la Figura 29, para el café Arábica lavado, se observa que la intensidad del color de la crema tiene una correlación negativa inversa con la cafeína en el perfil de temperatura ascendente. La textura de la crema presentó una relación positiva con los lípidos totales en el perfil ascendente y en el perfil plano con el índice color de la bebida. La intensidad del olor tuvo una relación inversa con el ácido clorogénico en el perfil ascendente, con el pH en el perfil plano y con la cafeína y los sólidos totales en el perfil descendente. La intensidad del tostado mostró una relación con el índice de la crema e inversa con la cafeína, ácido clorogénico y la viscosidad en el perfil de temperatura ascendente, en el perfil plano tuvo una relación negativa con el ácido

clorogénico y los sólidos totales mientras que en el perfil descendente tuvo una relación positiva con los lípidos. El cuerpo presentó una relación positiva con el índice color y la cafeína e inversa con el pH en el perfil plano, entre tanto, en el perfil de temperatura ascendente mostró una relación positiva con el índice color y viscosidad pero negativa con la cafeína, sólidos totales y los sólidos filtrados. El amargor se relacionó inversamente con los sólidos filtrados en el perfil plano, mientras que en el descendente mostró una relación positiva con los lípidos. Finalmente, la astringencia en el perfil de temperatura ascendente tuvo una relación positiva con la viscosidad, en el perfil descendente mostró una relación positiva con la cafeína y los sólidos filtrados mientras que fue negativa con el índice color y la viscosidad.

Como se puede ver en las Figuras 27-29, cada tipo de café es influido de manera diferente por los perfiles de temperatura, por lo que la relación entre las variables fisicoquímicas y los descriptores organolépticos de la bebida fueron difíciles de interpretar en su conjunto. Por este motivo, a continuación se realizaron los ACP solo para los descriptores organolépticos que permitirán analizar a través de dichos descriptores las características comunes entre los diferentes tipos de café en los perfiles de temperatura.

4.3.3 ANÁLISIS DE COMPONENTES PRINCIPALES – ACP POR PERFIL DE TEMPERATURA

A continuación se muestran los resultados del ACP por perfil de temperatura de extracción empleando las variables de los descriptores organolépticos. Cada una de las matrices de correlación fue evaluada mediante las pruebas estadísticas del índice de esfericidad de Bartlett y del índice de Kaiser-Meyer-Olkin KMO, que se muestran en la Tabla 18.

Estas pruebas indicaron que existe una estructura subyacente entre las variables por lo que se concluyó que los ACP son procedentes. (Pérez & Medrano, 2010).

Tabla 18. Índice de Bartlett e índice de Kaiser-Meyer-Olkin- KMO.

Perfil de temperatura	Pruebas			
	Esfericidad de Bartlett			KMO
	Razón n/p	Chi ² ₍₂₈₎	P valor	
Ascendente 88-93 °C	3,25	88,42	3,49E-08	,539
Plana 90-90 °C	3,25	75,43	3,13E-06	,646
Descendente 93-88 °C	3,25	65,99	6,63E-05	,504

Como se evidencia en la Tabla 17, con los resultados de la prueba KMO se duda de continuar con los análisis, sin embargo con la prueba de Bartlett a un nivel de significancia del 0.05 se rechaza la hipótesis nula de esfericidad. Las distribuciones chi-cuadrado asociadas a cada perfil de temperatura fueron asintóticas y suponen la normalidad multivariante de los datos. Por lo que no se duda de la normalidad conjunta, además algunos autores advierten de que la razón n/p debe ser menor que 5 para proceder con el ACP también se cumple.

En la Tabla 19 se muestra la proporción de variabilidad explicada por componente en cada uno de los perfiles de temperatura. La selección del número de factores se basó en el criterio del histograma de valores propios donde se seleccionan aquellos que son mayores o iguales a uno.

Ahora, teniendo en cuenta las comunalidades encontradas en cada uno de los perfiles de temperatura que se muestran en la Tabla 20, se observan que la cantidad de varianza con que contribuye cada variable a la solución final están por encima del 67%, lo que indica que se encuentran bien representadas sobre el primer y segundo plano.

Tabla 19. Proporción de variabilidad explicada por componente.

Componentes	Ascendente (88-93 °C)		
	Valores propios	% de la varianza	% Varianza acumulada
1	2,835	35,441	35,441
2	2,163	27,041	62,483
3	1,080	13,497	75,979
4	,821	10,265	86,245
Componente	Plana (88-93 °C)		
	Valores propios	% de la varianza	% Varianza acumulada
1	3,212	40,152	40,152
2	1,589	19,868	60,019
3	1,137	14,211	74,230
4	,769	9,617	83,848
Componente	Descendente (93-88 °C)		
	Valores propios	% de la varianza	% Varianza acumulada
1	2,691	33,638	33,638
2	1,952	24,402	58,040
3	,961	12,006	70,046
4	,898	11,224	81,271

Tabla 20. Comunalidades por perfiles de temperatura.

Variables	Ascendente (88-93 °C)		Plana (88-93 °C)		Descendente (93-88 °C)	
	Inicial	Extracción	Inicial	Extracción	Inicial	Extracción
Intensidad Color crema	1,000	,730	1,000	,697	1,000	,703
Textura Crema	1,000	,698	1,000	,782	1,000	,569
Intensidad Olor	1,000	,738	1,000	,766	1,000	,795
Intensidad Tostado	1,000	,782	1,000	,811	1,000	,641
Cuerpo	1,000	,901	1,000	,825	1,000	,709
Ácidoz	1,000	,471	1,000	,385	1,000	,701
Amargo	1,000	,880	1,000	,789	1,000	,726
Astringencia	1,000	,877	1,000	,883	1,000	,761

En las Figuras 30 - 32, muestran el primer plano factorial para los perfiles de temperatura que está conformado por el primer y segundo factor. Este plano factorial permite interpretar: 1) las correlaciones entre las variables originales y las

componentes que se encuentran delimitadas por el círculo negro, 2) las puntuaciones de las observaciones en las componentes y 3) las asociaciones entre las variables conformadas por los descriptores organolépticos y las observaciones conformadas por el tipo de café.

En cada uno de los gráficos, los puntos corresponden a los pesos de cada observación en la componente principal 1 (PC1) y componente principal 2 (PC2), las flechas representan la correlación de las variables con PC1 y PC2, el círculo negro lógicamente es de radio 1 e indican la extensión máxima teórica de las flechas, las elipses de colores representan las observaciones por tipo de café.

PERFIL DE TEMPERATURA ASCENDENTE (88-93 °C)

En este ACP se decidió retener tres factores basados en el criterio del histograma de los valores propios, donde se explica el 75.9% de la variabilidad de la nube de individuos, sin embargo solo serán mostrados los dos primeros factores, es decir el primer plano factorial.

Tabla 21. Componentes principales en el perfil de temperatura ascendente.

Variables	Componentes Principales		
	PC1	PC2	PC3
Intensidad Color Crema	0,2169	-0,8218	-0,0896
Textura Crema	0,8135	-0,0218	-0,1885
Intensidad Olor	0,6729	-0,4716	0,2516
Intensidad Tostado	0,5268	-0,6386	0,3114
Cuerpo	0,1233	-0,4163	-0,8444
Acidez	0,5427	0,3174	-0,2746
Amargo	-0,6350	-0,6695	0,1697
Astringencia	-0,8266	-0,3673	-0,2423

Como se destaca en la Tabla 21, la componente principal 1 (PC1), tiene una correlación positiva con los descriptores textura de la crema, intensidad del olor, acidez e intensidad tostado, que hace destacar el sabor y olor de los cafés espresso mientras que el amargo y la astringencia fueron menos evidentes (correlación

negativa) para los catadores en los espressos realizados en este perfil de temperatura. En la componente principal 2 (PC2), se presentan correlaciones negativas altas con la intensidad del color de la crema y la intensidad de tostado. Esto hace referencia a los espressos que muestran un aspecto visual pobre y con poco gusto. Finalmente en la componente principal 3 (PC3), se destacan aquellos espressos que poseen un cuerpo pobre.

En la Figura 30, se evidencia que en este perfil de temperatura, las variables astringencia y amargo están relacionadas entre sí, pero de manera inversa con la acidez, por otro lado las variables intensidad de olor e intensidad de tostado se encuentran relacionadas entre ellas.

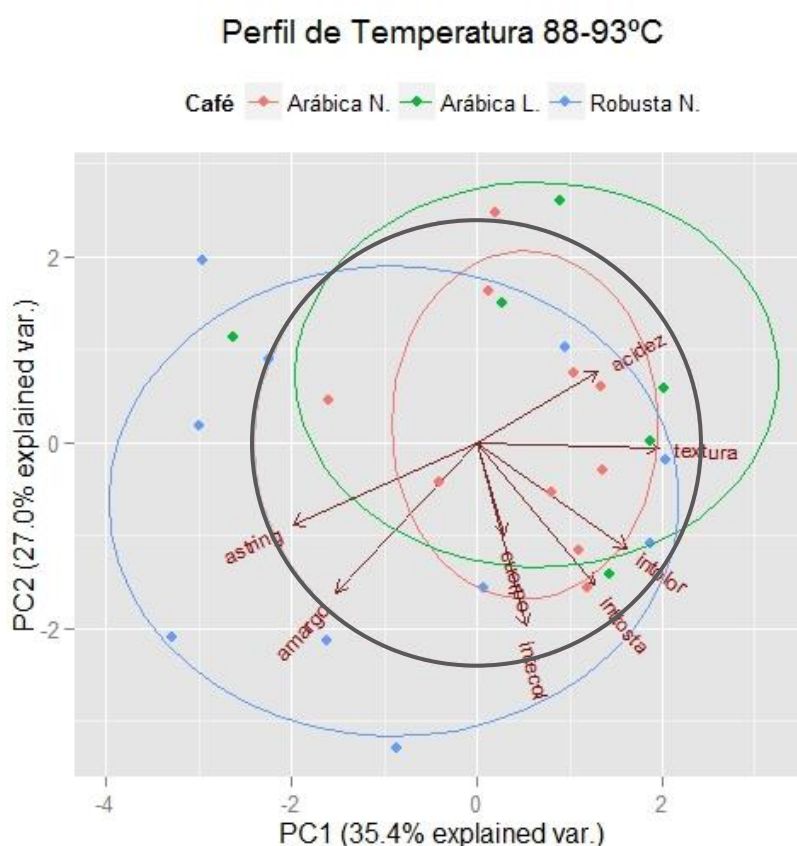


Figura 30. Primer plano factorial para el perfil de temperatura ascendente (88-93 °C).

Ahora, si se observan los tipos de café, se puede observar que los espressos preparados con Arábica Natural se describieron como de buena acidez pero baja textura de la crema, cuerpo e intensidad olor, tostado y color de la crema. El café Arábica Lavado presenta buena acidez, textura de la crema característica y baja intensidad de olor y cuerpo. El café Robusta Natural evidencia que los catadores lo han catalogado como un café con un ligero balance entre acidez - amargo y astringente, tendiendo más a los segundos, con buena acidez pero de baja intensidad de olor y tostado.

PERFIL DE TEMPERATURA PLANA (90-90 °C)

En la Tabla 22, que se presenta a continuación se observan las tres componentes principales que explican 74,23% de la variabilidad de los datos. En la Tabla se muestra que la PC1, se correlaciona positivamente con variables como la astringencia y negativamente con variables como textura de la crema, intensidad del olor y cuerpo, definiendo esta componente como de sabor astringente y de poco gusto. La PC2, se correlaciona positivamente con descriptores como intensidad del color de la crema y amargo, catalogando la PC2 como de sabor amargo y de buen aspecto físico. Por otro lado la PC3, se identifica con un tostado intenso.

Tabla 22. Componentes principales en el perfil de temperatura Plana.

Variables	Componentes Principales		
	PC1	PC2	PC3
Intensidad Color			
Crema	-0,0387	0,6809	-0,4815
Textura Crema	-0,7235	0,4718	0,1910
Intensidad Olor	-0,8179	0,1320	0,2827
Intensidad Tostado	-0,5035	0,1382	-0,7340
Cuerpo	-0,7571	0,3678	0,3407
Ácidoz	-0,4987	-0,3334	0,1598
Amargo	0,5135	0,6995	0,1895
Astringencia	0,8242	0,3620	0,2688

En la Figura 31, se evidencia que en este perfil de temperatura, la variable textura de

la crema y el cuerpo están correlacionados positivamente, así como las variables amargor y astringencia se relacionan positivamente pero que con la acidez la correlación es inversa. Entre la intensidad del tostado y la intensidad del olor igualmente existe una correlación positiva entre ellas.

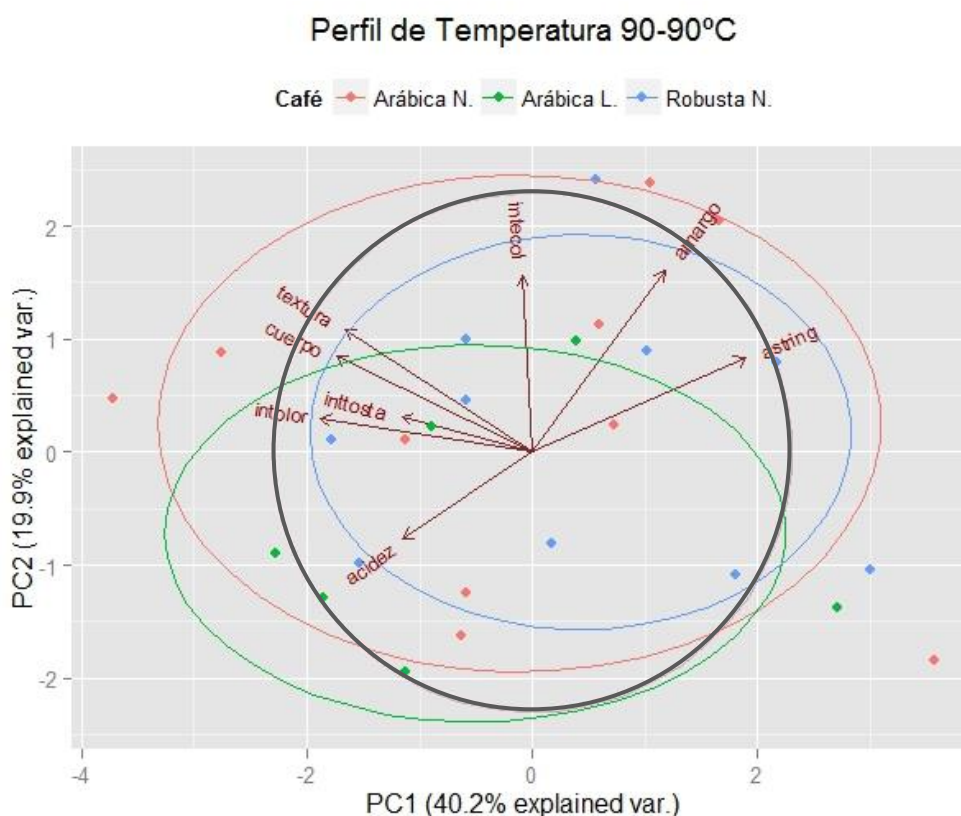


Figura 31. Primer plano factorial para el perfil de temperatura ascendente (90-90 °C).

En cuanto al tipo de café, se puede notar que los espressos hechos con café Arábica Natural realzan descriptores organolépticos como el amargor y la astringencia, baja acidez, baja intensidad olor, cuerpo y textura de la crema. El café Arábica Lavado es representado por variables como menor acidez y menor intensidad del olor, cuerpo y textura de la crema. El café robusta Natural, se ve constituido por descriptores como amargor, astringencia con menos acidez cuerpo y textura de la crema.

PERFIL DE TEMPERATURA DESCENDENTE (93-88 °C)

En la Tabla 23, se exponen las 3 componentes principales que explican el 70% de la variabilidad de los datos. En ella se observa que la PC1, se correlaciona positivamente con descriptores como textura de crema, intensidad de olor y cuerpo y se asocia negativamente con la astringencia, lo que define a esta componente como sabor del espresso. En la PC2, la correlación negativa alta de la intensidad del color de la crema, intensidad del tostado y la acidez definen la componente como de aspecto visual y acidez. Por otro lado la PC3 presenta correlaciones muy bajas en todas las variables por lo que es difícil su interpretación.

Tabla 23. Componentes principales en el perfil de temperatura Descendente.

Variables	Componentes Principales		
	PC1	PC2	CP3
Intensidad Color Crema	-0,0799	-0,7966	0,2471
Textura Crema	0,7169	-0,1891	0,1381
Intensidad Olor	0,7188	-0,4860	-0,2048
Intensidad Tostado	-0,0578	-0,6568	-0,4540
Cuerpo	0,8033	0,1400	-0,2115
Ácidoz	0,1876	-0,6079	0,5442
Amargo	-0,5288	-0,4370	-0,5053
Astringencia	-0,8310	-0,1851	0,1897

En la Figura 32, se ve que la variable cuerpo se encuentra inversamente relacionada con el astringencia. La variable textura se encuentra muy correlacionada con la variable intensidad de olor. Entre el tostado y la intensidad del olor igualmente existe una correlación positiva.

Si se observan los tipos de café, los espressos elaborados con el Arábica Natural y Arábica Lavado, muestran un buen cuerpo, baja astringencia, amargor, acidez y textura de la crema, definiendo un espresso de acidez y sabor ligero y buen cuerpo. Finalmente, el café el Robusta Natural tiene un comportamiento donde hay un buen cuerpo y menor astringencia y amargor lo que expresa un espresso de sabor ligero.

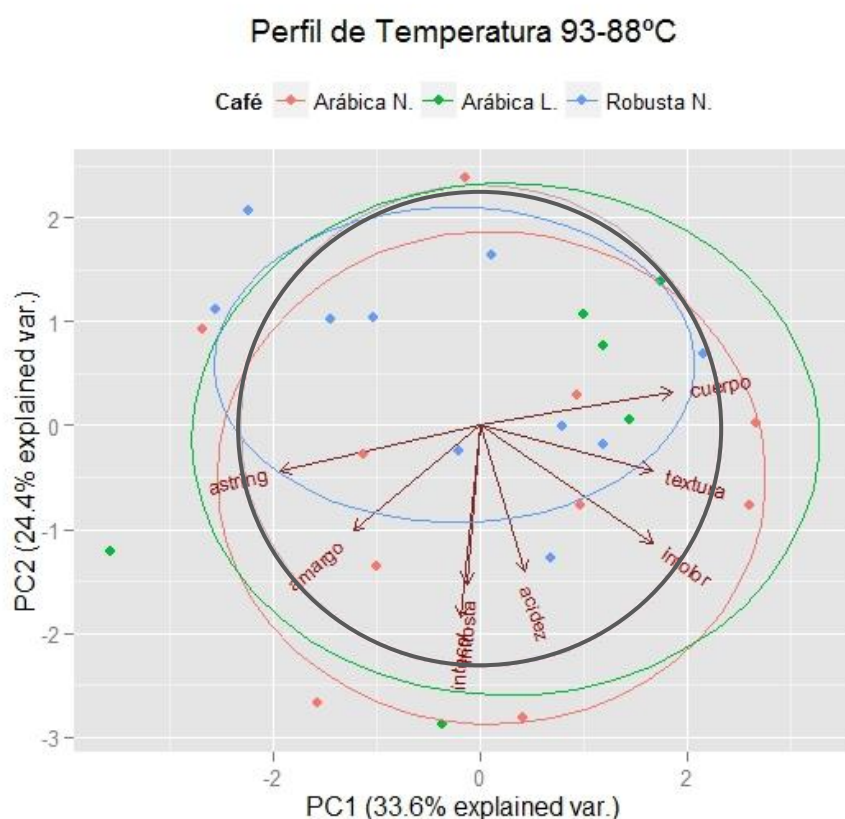


Figura 32. Primer plano factorial para el perfil de temperatura descendente (93-88 °C).

4.3.4. RELACIÓN DE LAS PROPIEDADES FÍSICOQUÍMICAS Y ORGANOLÉPTICAS CON LA TEMPERATURA DE EXTRACCIÓN

Las especies de café Arábica (*Coffea arabica* L.) y Robusta (*Coffea canephora* P.) presentan diferencias en la calidad y en perfiles sensoriales que requieren de adecuados métodos analíticos y paneles de degustación capacitados para cuantificar estas diferencias. (Fox, et al. 2013)

Según Parenti et al, (2014) la temperatura de extracción es un factor importante en la preparación de un café espresso por que se relaciona con los parámetros físicos y químicos. El contenido de sólidos totales y filtrados conllevan a percibir descriptores organolépticos como el amargor y la astringencia. Por ejemplo, cuando temperatura de extracción es alta hay un exceso de amargura y astringencia desagradable. Los

bajos valores del pH tuvieron una alta inusual percepción de la acidez que es determinada por la cafeína y los ácidos clorogénicos que conllevan a la acidez total, así como contribuyen a la astringencia y amargura. Por otro lado relaciona positivamente la densidad, la viscosidad y los lípidos del café con el cuerpo y la crema. Aunque la densidad en este estudio no fue significativamente diferente.

Finalmente, este estudio demostró que el uso de los perfiles de temperatura en el proceso de extracción de los espresso, es un factor clave para determinar la calidad de cualquier tipo de café, modificando los descriptores organolépticos al punto de poder obtener un espresso con características deseadas. Esto es interesante desde el punto de vista comercial, ya que se podrán potencializar aquellos cafés que no son tan apetecidos en el mercado.

5. CONCLUSIONES

1. Se utilizaron tres técnicas de imputación de datos Listwise o case deletion CD, Imputación Múltiple – MI y la Hot Deck – HD, de todas ellas, hasta el momento la más fiable es la MI, dada las bondades de la técnica aportada por esta y por otras investigaciones previas.
2. Mediante el uso del Análisis Factorial bayesiano a través MCMC se pudo construir el Índice de calidad del café mediante un modelo constituido por diferentes tipos de variables (continuas y ordinales). Los resultados obtenidos con el factor latente de las puntuaciones proporcionaron una menor dimensión que fue conveniente para resumir las variables observadas del estudio.
3. Con el cálculo de las cargas factoriales y el porcentaje de varianza explicada, se pudo observar que variables como almendra defectuosa, merma y almendra sana se destacan notablemente en el factor, así como para la almendra defectuosa, la varianza del error estimado implicó que casi toda su variabilidad puede ser explicada solo por el factor latente, débilmente también se podrían tener en cuenta la merma, olor el color y la almendra sana.
4. Mediante el índice de calidad hallado se pudo zonificar la calidad del café en el departamento del valle del Cauca, dando resultados aceptables, sin embargo, cabe anotar que esto es una primera aproximación a la zonificación a través de las pruebas de taza y su relación con las propiedades agronómicas y organolépticas utilizando este tipo de análisis factorial

bayesiano.

5. Se adaptaron diferentes técnicas analíticas para la determinación de las propiedades fisicoquímicas más importantes del café espresso.
6. La modificación del tamaño de molido y los perfiles de temperatura usando un mismo tipo de café tostado, permite obtener una extracción en mayor o menor grado de los compuestos del café a lo largo del tiempo de erogación, dando como resultado espressos con características muy distintas.
7. El espresso producido con café Robusta Natural usando el perfil de temperatura ascendente presenta cuerpo ligero y sabor fuerte, mientras que con la curva descendente se obtiene un espresso con buen índice de crema, cuerpo ligero, menor acidez y sabor agradable.
8. El espresso producido con café Arábica Natural usando el perfil de temperatura ascendente revelo un mayor cuerpo, sabor más intenso y balanceado si se compara con el obtenido en el perfil de extracción plana. Con el perfil de temperatura descendente se obtiene un espresso de buen cuerpo y sabor más ligero y noble comparado con los espressos obtenidos mediante perfil ascendente y extracción plana.
9. Los espressos obtenidos a partir de Arábica Lavado usando el perfil de temperatura ascendente se describe como de cuerpo ligero y sabor intenso, pero agradable, con balance entre ácidos-amargos, con tendencia hacia amargos comparados con los que se obtuvieron con el perfil de temperatura plana. Usando la curva de temperatura descendente, los espressos se pueden describir como menos intensos, con una suave acidez y un cuerpo notable que le confieren una gran cantidad de sensaciones organolépticas.

10. La relación entre las características fisicoquímicas como pH, color, polifenoles, cafeína, lípidos, ácido clorogénico, índice crema, turbidez, sólidos totales, sólidos filtrados, viscosidad y los descriptores organolépticos como intensidad color, textura, intensidad olor, intensidad tostado, cuerpo, acidez, amargo y astringencia; contribuyen según los resultados obtenidos a la interpretación de la calidad del café espresso elaborado con diferentes tipos de café y perfiles de temperatura.

11. Este estudio demuestra que la utilización de los perfiles o curvas de temperatura con la máquina de café RANCILIO XCelsius permite modificar las propiedades fisicoquímicas y exaltar o modular los descriptores organolépticos de un espresso.

BIBLIOGRAFIA

ABAYOMI K., GELMAN A., LEVY M. Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. Appl. Statist. 2008. 57(3): 273–291.

ALBANESE D., DI MATTEO M., POIANA M., SPAGNAMUSSO S. Espresso coffee (EC) by POD: Study of thermal profile during extraction process and influence of water temperature on chemical–physical and sensorial properties. *Food Research International*. 2009. (42):727–732.

ALFARO R., FUENZALIDA M. Nota Técnica: Imputación Múltiple en Encuestas Microeconómicas. *Cuadernos De Economía*. Banco Central De Chile - Columbia University. 2009. (46): 273-288.

ALLISON, P.D. *Missing Data*. Sage University. Papers Series on Quantitative Applications in the Social Sciences Series 07 -136. 2002.

ALVARADO A., PUERTA G. La variedad Colombia y sus características de calidad física y en taza. *Avances Técnicos Cenicafé*. 2002. (303):1-4.

AMON I. Guía metodológica para la selección de técnicas de depuración de datos. Universidad Nacional De Colombia. Facultad De Minas. Escuela De Sistemas. 2010.

ANDERSON B., SHIMONI E., LIARDON R., LABUZA T. The diffusion kinetics of carbon dioxide in fresh roasted and ground coffee. *J. Food Eng.* 2003a. (59): 71–78.

ANDERSON T. *An introduction to multivariate statistical analysis*. 3ed, New York,

USA. John Wiley. 2003b.

ANDUEZA S., MAEZTU L., DEAN B., DE PEÑA M., BELLO J., CID C. Influence of water pressure on the final quality of Arábica espresso coffee. Application of multivariate analysis, J Agric Food Chem. 2002. (50):7426–7431.

ANDUEZA S., DE PEÑA M., CID C. Chemical and sensorial characteristics of espresso coffee as affected by grinding and torrefacto roast. J Agric Food Chem. 2003a. 51(24):7034–7039.

ANDUEZA S., MAEZTU L., PASCUAL L., IBAÑEZ C., DE PEÑA M., CID C., Influence of extraction temperature on the final quality of espresso coffee. J Sci Food Agric. 2003b. 83(3):240–248.

ANDUEZA S., VILA M^a., PEÑA M^a., CID C. Influence of coffee/water ration the final quality of espresso coffee. J Agric Food Chem. 2007. (87):586–592.

ARCILA J., FARFÁN F., MORENO A., SALAZAR L., HINCAPIÉ E. Sistemas de producción de café en Colombia. Prácticas de cultivo. Caficultura Manejo de cafetales. Cenicafé. 2007.

ÁVILA C. Una aplicación del procedimiento Hot Deck como método de imputación. Capítulo III. Método de imputación hot deck .Trabajo Monográfico Para optar el Título Profesional de Licenciado en Estadística. Universidad Mayor de San Marcos. 2002.

AZUR M., STUART E., CONSTANTINE F., LEAF P. Multiple imputation by chained equations: What is it and how does it work? International Journal of Methods in Psychiatric Research. 2011. 20(1):40-49.

BADLER C., ALSINA S., PUIGSUBIRÁ C., VITELLESCHI M. Tratamiento De Bases De Datos Con Información Faltante Según Análisis De Las Pérdidas Con Spss. Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de

Estadística (IITAE). Novenas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística. 2004.

BAGGENSTOSS J., POISSON L., KAEGI R., PERREN R., ESCHER F. Coffee Roasting and Aroma Formation: Application of Different Time-temperature Conditions. *Journal of Agricultural and Food Chemistry*. 2008. 56(14), 5836-5846.

BAZZARA F. The espresso coffee production system Gorizia: Edizioni 2000. 2004.

BARON F., TELLEZ F. Capítulo 5: Diferencias que presenta una variable numérica entre varios grupos. *Apuntes De Bioestadística*. 2011.
<http://www.bioestadística.uma.es/baron/apuntes/ficheros/cap05.pdf>

BENÍTEZ C., DÍAZ G., MUÑOZ I. Tesis De Grado "Propuesta De Un Plan Estratégico De Comercialización Del Café Nacional Para Los Productores Del Departamento De San Miguel". Facultad De Ciencias Económicas. Universidad De Oriente. San Miguel. El Salvador. 2003.

BOENEKE C. MCGREGOR J. ARYANA K. The effect of sweeteners on the acceptability of dairy-based espresso beverages. *International Journal of Dairy Technology*. 2006. 59(1): 12-17.

CAPEL J., WICHMANN M., FERNANDEZ R., LIZARRAGA M., PÉREZ J., RIOBÓ P. *Café y estilo de vida saludable*. Centro de Información Café y Salud (CICAS) Ed. Editores Médicos, S.A. EDIMSA. 2010.

CAPPUCCI R., CATTANEO G., ERBACCI G., JOCHER U., A parallel implementation of a cellular automata based model for coffee percolation. *Parallel Comput*. 2001. 27(5): 685–717.

CARTER L. Solutions Missing Data in Structural Equation Modeling. *Research & Practice in Assessment*. 2006. 1(1).1-6.

CATANIA C., AVAGNINA M., Implicancias organolépticas de los polifenoles del vino. Curso de degustación de vinos. Instituto Nacional de Tecnología Agropecuaria Argentina Mendoza. 2007.

CHARURIN P., JAMES J., DEL CASTILLO M., Antioxidant activity of coffee model systems. J Agric Food Chem. 2002. (50):3751–3756.

CLARK J., MACRAE R. Coffee: vol 5. Related Beverages. Technology, London. Elsevier Applied Science. 1987.

CAMMENGA H., ZIELASKO B. Kinetics and development of boiler scale formation in commercial coffee brewing machines, ASIC. In Proceedings 17th colloquium. 1997. 284–289.

CONTI G, FRÜHWIRTH-SCHNATTER S, HECKMANN J, PIATEK R. Bayesian exploratory factor analysis. Journal of Econometrics 2014. (183) 31–57.

CROS J., GUYOT B., VINCENT J. Profil chromatographique de la fraction volatile du café. Différence entre cafés verts sains et puants influence de la torréfaction sur le grain et la boisson. J. Café Cacao Thé. 1979. 23(3):193-202.

CUADRAS S. La cata de café espresso. Revista fórum del café. 2009. http://www.forumdelcafe.com/pdf/F-36_Caf%C3%A9_espresso.pdf

CASTILLO D., AMES J., GORDON M. Effect of Roasting on the Antioxidant Activity of coffee Brews. J. Agric. Food Chem. 2002. (50): 3698–3703.

DONNER A. The relative effectiveness of procedures commonly used in multiple regression analysis for dealing with missing values. The American Statistician, 1982. (36):378-381.

DUARTE R., SUAREZ M., MORENO E., ORTIZ P. Análisis multivariado por

componentes principales de cafés tostados y molidos adulterados con cereales. *Cenicafe*. 1996. 47(2): 65–76.

DURRANT B. Imputation Methods for Handling Item-Non-response in the Social Science: A Methodological Review. ESRC National Centre for Research Methods and Southampton Statistical Science Research Institute. University of Southampton. NCRM Methods Review Papers. 2005.

ESPINAL C., MARTÍNEZ H., ACEVEDO X. La mirada del café en Colombia, Una mirada global de su estructura y dinámica 1991 – 2005. Documento de trabajo N° 59, Ministerio de Agricultura y Desarrollo Rural. Observatorio Agrociudades Colombia. 2005.

FEDERACIÓN NACIONAL DE CAFETEROS DE COLOMBIA – FNC. Valor agregado: Del árbol a la taza. Informe del Gerente General al LXVI Congreso Nacional de Cafeteros Federación: Permanencia, Sostenibilidad y Futuro. 2006.

FEDERACIÓN NACIONAL DE CAFETEROS DE COLOMBIA – FNC. Informe Comité Departamental de Cafeteros del Valle del Cauca. 2012-2013.

FERMÍN N., GALÁN H., GARCÍA J., BRACHO N. Evaluación de la calidad físicoquímica y sensorial de tres marcas comerciales de café tostado y molido. *Revista Científica UDO Agrícola*. 2012. 2(2): 428-438.

FOX G., WU A., YIRAN L. FORCE L. Variation in Caffeine Concentration in Single Coffee Beans. *J. Agric. Food Chem.* 2013. 61: 10772–10778

GARDNER D. Effect of certain ion combinations commonly found in potable water on rate of filtration through roasted and ground coffee. *Food Research*. 1958. (23): 76–84.

GELMAN A., RUBIN D. "Inference from Iterative Simulation Using Multiple

- Sequences (with discussion)." *Statistical Science*. 1992. 7(4):457-511.
- GOICOECHEA P. Imputación basada en árboles de clasificación. *EUSTAT*. 2002.
- GRAHAM J. Missing data analysis: making it work in the real world. *Annual Review of Psychology*. DOI: 10,1146/ annurev. Psych. 58,1104t 05,ern085530. 2009. (60): 549–576.
- GRAJALES T. El Análisis factorial. 2000. Web: <http://tgrajales.net/estfactorial.pdf>
- GUEVARA A., CASTAÑO J. Caracterización granulométrica del café colombiano tostado y molido. *Cenicafé*. 2005. 56(1):5-18.
- HAGERMAN A., RIEDL K., ALEXANDER J., SOVIK K., RITCHARD N., HARTZFELD P., RIECHEL T. High molecular weight plant polyphenolics (tannins) as biological antioxidants. *J. Agric. Food Chem*. 1998. (46):1887-1892.
- HEIDELBERGER P., WELCH P. Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*. 1983. 31 (6):1109-1144.
- HIROSE K., KAWANO S., KONISHI S. Bayesian Factor Analysis And Information Criterion. *Bulletin of Informatics and Cybernetics*. 2008. (40)
- HUMPHRIES D. Understanding Acidity in Coffee (The Goggles Do Nothing!). *Professional coffee ninja*. New York City. 2008.
<https://www.aikanekonacoffee.com/kona-coffee/appellations/>.
- ILLY E. The complexity of coffee, *Sci, Am*. 2002. (6): 92–98.
- ILLY A., VIANI R. Espresso coffee: the science of quality. 2nd ed. Elsevier Academic Press. London. 2005.
- ILLY E., NAVARINI L. The Espresso Coffee Foam. *Food Biophysics*. 2011.

(6):335–348.

INSTITUTO INTERAMERICANO DE COOPERACIÓN PARA LA AGRICULTURA - IICA. Protocolo de análisis de calidad del café. Programa cooperativo regional para el desarrollo tecnológico y modernización de la caficultura. Promecafe. Guatemala. 2010.

ISTITUTO NAZIONALE ESPRESSO ITALIANO. "Espresso Italiano Certificato". 2013.

KLINE R. Principles and Practice of Structural Equation Modeling. The Guilford Press. 2011.

KUEHL R. Diseño de experimentos. Principios estadísticos de diseño y análisis de investigación. Mexico: Thompson Learning. 2001.

LITTLE J., RUBIN D. Statistical Analysis with missing data. New York: Wiley. 1987.

MAGNANI M. Techniques for Dealing with Missing Data in Knowledge Discovery Tasks. University of Bologna. Department of Computer Science. Bologna – ITALY. 2004.

MARIN C., PUERTA G., Contenido de ácidos clorogénicos en granos de Coffea Arábica L. y Coffea Canephora según el desarrollo del fruto. Cenicafé. 2008. 59(1):07-28.

MARÍN S., ARCILA J., MONTOYA E., OLIVEROS C., Relación entre el estado de madurez del fruto del café y las características de beneficio rendimiento y calidad de la bebida. Cenicafé. 2003. 54(4):297-315.

MARTÍNEZ E. Análisis factorial. Apuntes para el curso de Métodos Multivariantes. Curso 2006 – 2010. Universidad de Antofagasta. 2008.

MARTINEZ M. La cata de café. Fórum del café.2013

http://www.forumdelcafe.com/pdf/F-35_La_cata_de_caf%C3%A9.pdf

MARWALA T. Computational Intelligence for Missing Data Imputation. Estimation and Management: Knowledge Optimization TECHNIQUES. University of Witwatersrand. South Africa. 2009.

MASELLA P., PGUERRINI L., SPINELLI S., CALAMAI L, SPUGNOLI P, ILLY F., PARENTI A. A new espresso brewing method. Journal of Food Engineering 2015. (146): 204–208

MEDIAVILLA M. Las becas y la continuidad escolar en el nivel secundario post-obligatorio. Un análisis de sensibilidad a partir de la aplicación del propensity Propensity Score Matching. 2010. (5):561-582.

MEDINA F., GALVÁN M. Imputación de datos: teoría y práctica. Serie 54. Estudios estadísticos y prospectivos. División de Estadística y Proyecciones Económicas. Naciones Unidas. 2007.

MEDINA J., RIAÑO C. Evaluación del rendimiento de extracción de algunas cafeteras. Cenicafé. 2006. 57 (1): 31-36.

MOLINERO L. Análisis de la varianza. Asociación de la sociedad española de hipertensión. 2003 <http://www.seh-lelha.org/pdf/anova.pdf>

MORA I., RODRÍGUEZ A. Comportamiento de la cafeína en un espresso. REVISTA DEL FÓRUM CAFÉ N°42. 2010.

MORALES P. “El Análisis Factorial en la construcción e interpretación de tests, escalas y cuestionarios Universidad Pontificia Comillas. Madrid Facultad de Ciencias Humanas y Sociales.2013.

MUÑOZ A. Diseño De Una Guía De Selección Del Sello Para Cafés Sostenibles. Universidad Tecnológica De Pereira. Facultad De Ciencias Ambientales. Administración Del Medio Ambiente. Pereira. 2007.

NAVARINI L., RIVETTI D. Water quality for espresso coffee. Food. Chem.2010. (122): 424–428.

NICOLI C., ANESE M., MANZOCCO L., LERICI R. Antioxidant properties of coffee brews in relation to the roasting degree. Food Science and Technology. 1997. 30(39): 292-297.

NUNES F., COIMBRA M., DUARTE C., DELGADILLO I. “Foam Ability, Foam Stability and Chemical Composition of Espresso Coffee as Affected by the Degree of Roas. J, Agric, Food Chem. 1997. (45): 3238-3243.

OIC- International Coffee Organization. Botanical Aspects. 2010.

OTERO D. Imputación de datos faltantes en un Sistema de Información sobre Conductas de Riesgo. Master en Técnicas Estadísticas. Universidad da Coruña. 2011.

PANGBORN. R., TRABUE M., LITTLE A. Analysis of coffee, tea and artificially flavoured drinks prepared from mineralized waters. Journal of Food Science. 1971. (36): 355–362.

PARENTI A., GUERRINI I., MASELLA P., SPINELLI S., CALAMAI L., SPUGNOLI P. Comparison of espresso coffee brewing techniques. Journal of Food Engineering 2014 (121): 112–117.

PÉREZ C. Técnica de análisis multivalente de datos. Aplicaciones con SPSS. Madrid: Pearson. 2004.

PÉREZ C., SANTÍN D. Minería de datos. Técnicas y Herramientas. ISBN: 8497324922. 2007.

PÉREZ E., MEDRANO L. Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. Revista Argentina de Ciencias del Comportamiento, 2010. 2(1): 58-66.

PUERTA G. La calidad de las variedades de café Coffea Arábica L. cultivadas en Colombia. Cenicafé. 1998. 49(4): 265-278.

PUERTA G. Influencia del proceso de beneficio en la calidad del café. Cenicafé 1999. 50(1):78-88.

PUERTA G. Influencia de los granos de café cosechados verdes, en la calidad física y organoléptica de la bebida. Cenicafé. 2000a. 51(2):136-150.

PUERTA G. Calidad en taza de algunas mezclas de variedades de café de la especie Coffea Arábica L. Cenicafé. 2000b. 51(1): 5-19.

PUERTA G. Riesgos para la calidad y la inocuidad del café secado. Cenicafé. Avance técnico 371. ISSN: 0120-0178. 2008.

PUERTAS G. "Factores de origen y proceso en la calidad y la química del café". Revista: Simposio Agroalimentario. Universidad De Córdoba Facultad De Ciencias Agrícolas. 2011a.

PUERTA G. "Composición química de una taza de café". Avances Técnicos Cenicafé ISSN: 0120-0178 ed: Blanecolor Ltda. 2011b. 414(414):1-12.

PUERTAS M., VILLEGAS P., ROJANO B., PELAEZ C. Comparación entre el estado de maduración del fruto de café (Coffee Arábica L.), el contenido de antocianinas y su capacidad antioxidante. Revista Cubana de Plantas Medicinales. 2012. 17(4): 360-367.

QUINN K. Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses. *Political Analysis*. 2004. (12): 338-353.

RAFTERY A., LEWIS S. "How Many Iterations in the Gibbs Sampler?" In *Bayesian Statistics 4* eds, J.M. Bernardo, J. Berger, A.P. Dawid and A.F Smith. Oxford: Oxford University Press.1992. 763- 773.

RAGHUNATHAN T., LEPKOWSKI J., VAN H. SOLENBEGGER P. Multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*. 2001. 27 (1): 85–95.

RATNAYAKE W., HOLLYWOOD R., O'GRADY E., STAVRIC B. Lipid content and composition of coffee brews prepared by different methods, *Food Chem Toxicol*. 1993. 31(4):263–269.

RIAÑO C. Efecto de la humedad del café crudo en las propiedades del café tostado. 2013. 9(2): 214-222.

RICHELLE M., TAVAZZI I., OFFORD E. Comparison of the antioxidant activity of commonly consumed polyphenolic beverages prepared per cup serving. *J Agric Food Chem*. 2001.49(7): 3438-42.

ROSA M., NICOLI, C., LERICI C. Quality characteristics of the espresso coffee in relation to the type of preparation (Italian). *Ind, Aliment*. 1986. (9), 629-633.

ROTH P., CAMPION, J., JONES S. The impact of four missing data techniques on validity estimates in human Resource management. *J. Bus. And Psych*. 1996. (11):101-116.

RUBIN D. "Inference and Missing Data" .*Biometrika*. 1976. (63): 581-592.

RUBIN D., SCHENKER N. "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse" *Journal of the American*

Statistical Association. 1986. (81):366-374.

RUBIN D. Multiple Imputation for Nonresponse in Surveys. New York. Wiley. 1987.

RUIZ M., UREÑA M., TORRES M., ESPINAL F. Los Mercados Del Café y De Los Cafés Especiales. Situación Actual Y Perspectivas. Programa Más Inversión Para El Desarrollo Alternativo Sostenible. MIDAS. 2009.

SCHAFER J. Analysis of Incomplete Multivariate Data. Chapman & Hall, London, 1997.

SCHAFER J., GRAHAM J. Missing Data. : Our View of the State of the Art. Psychological Methods. 2002. 7(2): 147-177.

SILVA L. BENAVIDES A. El enfoque bayesiano: otra manera de inferir Gaceta Sanitaria. 2001.15 (4): 341-346.

SINGLETON V., ROSSI J., Colorunetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. Amer. J. Enol. Viticult. 1965. (16):144-58.

SIVETZ, M. How acidity affects coffee flavour. Food Technology. 1972. (26):70-77.

SMITH L. A tutorial on Principal Components Analysis. Technical report. Univ. Of Otago.2002.

SPEER K., KÖLLING I. The lipid fraction of the coffee bean. J. Plant Physiol. Braz. J. Plant Physiol. 2006. 18(1):201-216.

STATACORP. Multiple-Imputation Reference Manual. Stata Press. 2009.

STUART E., AZUR M., FRANGAKIS C., LEAF P. Practical imputation with large

datasets: a case study of the Children's Mental Health Initiative. *American Journal of Epidemiology*. 2009. 169(9):1133–1139.

TERRADEZ J. Análisis de la varianza (ANOVA). Proyecto E-MATH. Uso de las TIC en asignaturas cuantitativas aplicadas. Universitat Oberta de Catalunya. 2003. <http://www.uoc.edu/in3/emath/docs/ANOVA.pdf>

TOBASURA I. La Crisis Cafetera. Una oportunidad para el cambio en las regiones cafeteras de colombia. *Revista agronomía ed*: 2006. 13(2): 5 – 17.

TRUGO L., MACRAE R. Chlorogenic acid composition of instant coffees. *Analyst*. 1984. 3(109): 263-266.

USECHE L., MESA D. Introducción A La Imputación De Valores Perdidos. Universidad Universidad Central de Venezuela. *Terra Nueva*. 2006 22(31): 127-151.

VAN S. Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical Methods in Medical Research*. 2007. (16):219–242.

VAN S., GROOTHUIS K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011. 45(3): 1-67.

VARELA J., BRAÑA T., GARCÍA A., RIAL B., VÁZQUEZ X, Estimación de la respuesta de los "no sabe/no contesta" en los estudios de intención de voto, Universidad de A Coruña - Universidad de Santiago .*REIS - Revista Española de Investigaciones Sociológicas*. 1998. (83): 269-287.

WANG N. Physicochemical Changes of Coffee Beans During Roasting. In partial fulfilment of requirements for the degree of Master of Science. *Food Science*. The University of Guelph, Ontario, Canada. 2012.