



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Artes en Estadística Aplicada

**MODELO DE REGRESIÓN LINEAL MULTIVARIADO APLICADO A LA
RELACIÓN ENTRE VARIABLES QUÍMICAS DEL CLINKER Y VARIABLES
FÍSICAS DEL CEMENTO.**

Diego Enrique Valle García

Asesorado por el Mtro. Edgar Saúl Loaiza Salguero

Guatemala, junio de 2018

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**MODELO DE REGRESIÓN LINEAL MULTIVARIADO APLICADO A LA
RELACIÓN ENTRE VARIABLES QUÍMICAS DEL CLINKER Y VARIABLES
FÍSICAS DEL CEMENTO.**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

DIEGO ENRIQUE VALLE GARCÍA

ASESORADO POR EL MTRO. EDGAR SAÚL LOAIZA SALGUERO

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, JUNIO DE 2018

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. Ángel Roberto Sic García
VOCAL II	Ing. Pablo Christian de León Rodríguez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Oscar Humberto Galicia Nuñez
VOCAL V	Br. Carlos Enrique Gómez Donis
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADORA	Dra. Mayra Virginia Castillo Montes
EXAMINADOR	Mtro. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. Juan Carlos Fuentes Montepeque
SECRETARIA	Inga. Lesbia Magalí Herrera López.

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

MODELO DE REGRESIÓN LINEAL MULTIVARIADO APLICADO A LA RELACIÓN ENTRE VARIABLES QUÍMICAS DEL CLINKER Y VARIABLES FÍSICAS DEL CEMENTO.

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 03 de julio de 2017.

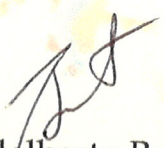
Diego Enrique Valle García

Ref.APT-2018-014

Como Coordinador de la Maestría en Artes en Estadística Aplicada del Trabajo de Graduación titulado **“MODELO DE REGRESIÓN LINEAL MULTIVARIADO APLICADO A LA RELACIÓN VARIABLES QUÍMICAS DEL CLINKER Y VARIABLES FÍSICA DEL CEMENTO”** presentado por el Ingeniero Químico **Diego Enrique Valle García**, apruebo y recomiendo la autorización del mismo.

Atentamente,

“Id y Enseñad a Todos”


MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador(a) de Maestría
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala



Guatemala, junio de 2018.

Cc: archivo/L.Z.L.A.

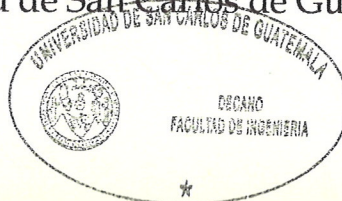
Ref.APT-2018-014

El Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen y dar el visto bueno del revisor y la aprobación del área de Lingüística al Trabajo de Graduación titulado **"MODELO DE REGRESIÓN LINEAL MULTIVARIADO APLICADO A LA RELACIÓN VARIABLES QUÍMICAS DEL CLINKER Y VARIABLES FÍSICA DEL CEMENTO"** presentado por el Ingeniero Químico **Diego Enrique Valle García**, correspondiente al programa de Maestría en Artes en Estadística Aplicada; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

MSc. Ing. Murphy Olympto Paiz Recinos
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala



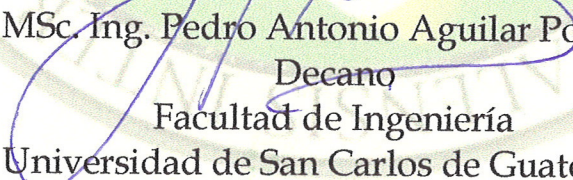
Guatemala, junio de 2018.

Cc archivo/L.Z.L.A.

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Postgrado, al Trabajo de Graduación de la Maestría en Artes en Estadística Aplicada titulado: **"MODELO DE REGRESIÓN LINEAL MULTIVARIADO APLICADO A LA RELACIÓN VARIABLES QUÍMICAS DEL CLINKER Y VARIABLES FÍSICA DEL CEMENTO "** presentado por el Ingeniero Químico **Diego Enrique Valle García**, procede a la autorización para la impresión del mismo.

IMPRÍMASE.

"Id y Enseñad a Todos"


MSc. Ing. Pedro Antonio Aguilar Polanco
Decano
Facultad de Ingeniería
Universidad de San Carlos de Guatemala



Guatemala, junio de 2018.

Cc archivo/L.Z.L.A.

ACTO QUE DEDICO A:

- Dios** Por ser mi principal fortaleza y regalarme la dicha de concluir exitosamente este logro.
- La Virgen María** Por ser mi apoyo espiritual, a quien siempre he buscado para lograr lo que me propongo.
- Mis padres** Antonio Valle y Cecilia García, porque sin ellos definitivamente esto no habría sido posible, pues son quienes me han formado en la vida.
- Mis hermanas** Mónica y Mariana, porque a pesar de todo siempre han sido un apoyo incondicional.
- Mi abuela** Margarita Flores (q.e.p.d.), por ser mi motivación para lograr cada uno de mis objetivos y siempre haberme acompañado con sabios consejos.
- Amigos** Diego Romero, Cristhian Monroy, Luis Mármol, José Rebelo, María Fernanda Cuéllar, Lucy Herrarte y Sergio Ordoñez, por haber formado parte de este proceso al estar presentes para siempre dar una palabra de apoyo.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser mi casa de estudios nuevamente en la que me he formado como profesional.
Facultad de Ingeniería	Por darme la oportunidad de seguir formando mi carrera como ingeniero y darme las herramientas necesarias para lograr mis metas.
Cementos Progreso	Por ser el lugar que me abrió las puertas para realizar mi estudio y me ofreció todos los recursos necesarios para culminarlo.
Mi padrino	Por ser una persona que representa un ejemplo a seguir y por compartirme consejos para ser una mejor persona y profesional
Doctora Mayra Castillo	Por todo su esfuerzo puesto en el programa de maestría y tomarse el tiempo para apoyar a cada uno de nosotros, para lograr nuestro objetivo en este programa de maestría.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES	III
LISTA DE SÍMBOLOS	V
GLOSARIO.....	VVII
RESUMEN	IX
PLANTEAMIENTO DEL PROBLEMA	XI
OBJETIVOS	XVII
RESUMEN DEL MARCO METODOLÓGICO	XIX
INTRODUCCIÓN	XV
1. MARCO REFERENCIAL.....	1
1.1. Estudios previos	1
2. MARCO TEÓRICO	9
2.1. Clinker y cemento.....	9
2.1.1. Procesos de producción del cemento	9
2.1.2. Clinker.....	10
2.1.3. Fases minerales del clinker	11
2.1.3.1. Alita (C_3S).....	11
2.1.3.2. Belita (C_2S)	11
2.1.3.3. Aluminato (C_3A)	11
2.1.3.4. Ferrita (C_4AF).....	12
2.1.3.5. Difracción de rayos X	12
2.1.4. Cemento.....	12
2.1.4.1. Resistencia a la compresión	13
2.1.4.2. Parámetros de calidad	13

2.2.	Fundamentos estadísticos	14
	2.2.1.Regresión múltiple	14
	2.2.2.Supuestos estadísticos de la regresión múltiple	14
	2.2.2.1. Independencia.....	14
	2.2.2.2. Homocedasticidad	15
	2.2.2.3. Normalidad	16
	2.2.2.4. Multicolinealidad.....	16
	2.2.2.5.Prueba de hipótesis	17
	2.2.3.Principio de parsimonia	21
3.	PRESENTACIÓN DE RESULTADOS	23
4.	DISCUSIÓN DE RESULTADOS.....	33
	CONCLUSIONES	41
	RECOMENDACIONES.....	43
	REFERENCIAS BIBLIOGRÁFICAS	45
	APÉNDICES.....	49

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Proceso de fabricación de cemento	10
2.	Q-Q plot de los residuos.....	28
3.	Prueba de Anderson para la belita (C_2S).....	30
4.	Prueba de Anderson para la alita (C_3S).....	31

TABLAS

I.	Definición de variables independientes y dependientes	XX
II.	Análisis de la varianza del modelo global propuesto	23
III.	Análisis de la varianza por variable del modelo propuesto.....	23
IV.	Análisis de contrastes globales del modelo propuesto	24
V.	Análisis de contrastes individuales del modelo propuesto.....	24
VI.	Índices de inflación de las fases minerales del clinker para la resistencia a la compresión a 28 días del cemento UGC	25
VII.	Jerarquía de fases minerales, según su impacto en R28d.....	26
VIII.	Determinación de coeficientes para modelo predictivo de la resistencia a la compresión del cemento a 28 días.....	26
IX.	Análisis de contrastes globales del modelo propuesto	27
X.	Prueba de Kolgomorov-Smirnov para normalidad de residuos	28
XI.	Análisis de multicolinealidad de las variables	29

LISTA DE SÍMBOLOS

Símbolo	Significado
C_2S	Abreviatura para fase mineral Belita
C_3A	Abreviatura para fase mineral Aluminato
C_3S	Abreviatura para fase mineral Alita
C_4AF	Abreviatura para fase mineral Ferrita
$R28d$	Resistencia a la compresión a 28 días
T	Valor T de la distribución T-Student
ϵ	Error
N	Población estudiada
n	Tamaño de muestra
y_i	Variable respuesta del modelo de regresión
x_i	Variable independiente del modelo de regresión
β_i	Coefficientes del modelo de regresión
σ	Desviación estándar para una población
σ^2	Varianza de la distribución
s	Desviación estándar muestral
\bar{x}	Media aritmética
H_a	Hipótesis alterna
H_o	Hipótesis nula
r^2	Coefficiente de determinación
Z	Valor Z de la distribución normal
F	Valor F de la distribución Fisher
k	Cantidad de desfases de una distribución normal
$V(\epsilon)$	Varianza del error

GLOSARIO

ANOVA	En estadística, por sus siglas en inglés representan el análisis de la varianza, en español es conocida como ANDEVA.
ASTM	Siglas en inglés para la American Society of Testing Materials.
Cemento UGC	Cemento de uso general en la construcción, también conocido como cemento tipo GU.
COGUANOR	Comisión Guatemalteca de Normas.
Cubix3D	Equipo para análisis por medio de rayos X a través de la técnica de difracción.
Difractograma	Es la representación gráfica obtenida de un difractómetro, a partir de la decodificación de conteos de intensidad de energía producida por la difracción de rayos x.
Holcim	Empresa suiza líder en la distribución de cemento y materiales de construcción con presencia en más de 70 países y que promueve innovaciones en la industria cementera.

InfoStat	Software para análisis estadístico de aplicación general desarrollado bajo la plataforma Windows.
PanalyticalEmpyrean	Equipo de análisis por rayos X que utiliza la técnica de difracción para la identificación y cuantificación de fases minerales en materiales sólidos.
Regresión	En estadística se refiere a la tendencia de una medición extrema a presentarse más cercana a la media en otra medición.
XRD	Siglas en inglés para definir la técnica de difracción de rayos X. En español es conocido como DXR.

RESUMEN

La industria cementera en la que se realizó la investigación cuenta con más de 100 años de experiencia en el mercado de la construcción y con los años ha podido implementar mejoras en el proceso de producción de cemento.

Esta investigación forma parte de estas metodologías de optimización, ya que se encontró una nueva oportunidad de mejora con la resistencia a la compresión a 28 días del cemento tipo UGC. Dado que esta propiedad es una de las más importantes, se buscó la relación que tiene con la concentración de fases minerales presentes en el clinker utilizado para su producción, porque con esta información es posible a tener un mejor control en la calidad.

Para llevar a cabo lo planteado en este estudio, se utilizó la técnica estadística de regresión lineal múltiple, con la que se analizó la influencia que cada uno de los minerales presentes en el clinker tiene sobre la propiedad física de resistencia a la compresión. Inicialmente se determinó que las cuatro fases minerales no pueden ser tomadas en cuenta para un modelo de regresión lineal porque, según el análisis de la varianza y un análisis de contrastes individuales, se observó colinealidad con algunas variables independientes. Posterior a esto se seleccionaron las fases de belita y alita, las que según el análisis de índice de inflación presentan la mayor influencia. Conocidas estas variables se procedió a la creación del modelo lineal de dos variables para utilizarlo como herramienta de predicción de la resistencia del cemento, modelo que fue validado, a partir del cumplimiento de los cuatro supuestos estadísticos para la regresión lineal múltiple: normalidad, homocedasticidad, independencia y multicolinealidad.

PLANTEAMIENTO DEL PROBLEMA

Contexto general

La presente investigación se llevó a cabo en una cementera guatemalteca, empresa líder en la producción de cemento a nivel nacional, que cuenta con más de 100 años como la productora principal del país de materiales para la construcción, principalmente de cemento tipo Pórtland con diferentes características, según la aplicación que se le quiera dar. Su principal producto es el cemento tipo UGC, de uso general en la construcción, el cual es producido tanto para cubrir las necesidades del territorio nacional como otros países americanos.

Uno de los pilares principales de la empresa es brindar a sus clientes un producto de calidad de clase mundial, para ello cuenta con estrictos controles de calidad establecidos, basados en normativas, tanto nacionales como internacionales, con los cuales se monitorea, reporta y corrige cualquier desviación detectada, ya sea en el producto final o algún subproducto del proceso. El departamento de control de calidad está dividido en dos grandes áreas de análisis; laboratorio de análisis químico y laboratorio de análisis físico, debido a que surge la necesidad de mantener condiciones específicas para cada tipo de análisis, con el fin de apegarse a lo establecido en las instrucciones internacionales del cemento.

El departamento de control de calidad de esta empresa, en su laboratorio, cuenta con una base de datos interna y confidencial amplíase las propiedades químicas y físicas tomadas para materias primas, producto intermedio y producto terminado de la producción de cemento, los cuales son recolectados y

analizados, con el fin de tener controlado todo el proceso de producción para evitar cualquier desviación en la fabricación del cemento y así cumplir con los estándares de calidad.

Los resultados obtenidos para cada una de las propiedades medidas presenta una tendencia propia y se analiza de manera individual cada una de ellas, basadas en rangos descritos en un plan de calidad apegado a normas internacionales para ofrecer un producto que cumpla con los requisitos nacionales e internacionales vigentes, estos valores son tomados por medio de rayos X, tanto por técnicas de fluorescencia como por difracción.

Las propiedades físicas del producto terminado, se vuelven prioritarias debido a que son críticas en cuanto a normativas, porque son aquellas que el cliente espera obtener para cumplir con sus necesidades. Para el cemento de tipo UGC, la propiedad que representa una mayor criticidad en cuanto a los parámetros de calidad es la resistencia a 28 días, que presenta los indicios del comportamiento del concreto una vez ha sido aplicada.

Descripción del problema

Durante la producción del cemento se debe tener un estricto control en el cumplimiento de los requisitos mínimos estandarizados por un plan de calidad. Actualmente se cuenta con una nueva metodología para la cuantificación de las fases minerales con las que cuenta la materia prima, para la fabricación del cemento, principalmente el clinker, y ésta es la técnica de difracción de rayos X, para determinar si éste cumple con las condiciones mínimas esperadas en su composición química. También se analiza el producto final, cemento, en su variable de resistencia a la compresión a 28 días,

igualmente para tener la certeza que el producto despachado cumple con lo que las normas nacionales dictan para ser distribuido.

Sin embargo, se desconoce cuál es efecto directo que la concentración de estas fases tiene específicamente sobre la resistencia a la compresión a 28 días del cemento, por lo que se requiere la implementación de una nueva metodología que relacione el análisis XRD y resistencias del cemento.

Determinar el grado de asociación que tienen estas fases minerales presentes en el clinker con el resultado de la resistencia a la compresión a 28 días del cemento, aportó información clave que fue utilizada para la optimización del proceso de producción, porque al contar con una predicción de las condiciones finales del cemento, a partir de la concentración mineralógica del clinker, costos y diferentes recursos pudieron ser reducidos con un producto final que cumple todos los requisitos mínimos esperados.

La Alita (C_3S), Belita (C_2S), Ferrita (C_4AF) y Aluminato (C_3A) son las fases minerales de mayor presencia en el clinker, y es importante la determinación de dichas fases minerales presentes en él por medio de difracción de rayos x, porque brinda información de vital importancia sobre la composición química con la que se está elaborando el producto que saldrá a la venta, pero no solo identificarlas es importante, sino establecen cuál o cuáles de todas estas fases tienen mayor impacto en la resistencia a la compresión a 28 días, para contar con un mejor monitoreo sobre la calidad del clinker que es utilizado para la fabricación de cemento.

Con la generación de un análisis estadístico de los datos generados durante un mes de producción, se determinó una correlación para llegar a un modelo de regresión que permitió establecer la relación que existe entre las

fases minerales presentes en los materiales utilizados para la fabricación de cemento con la resistencia que éste presente o determinar la ausencia de relación entre dichas fases con las características físicas presentadas por el producto terminado.

La aplicación de diferentes pruebas estadísticas a todos los datos y a la correlación establecida permitió dotar de confiabilidad a los resultados presentados, los cuales sirvieron como soporte al momento de establecer la influencia que cada una de las fases tuvo sobre la resistencia final del cemento, con ello se dotó de fundamento estadístico cada uno de los modelos presentados, o como puede darse el caso, así fundamentar concretamente que dicha relación supuesta entre cada variable estudiada, no existe.

Formulación del problema

Pregunta general

¿Cuál es el grado de asociación y qué impacto genera en la resistencia a la compresión del cemento tipo UGC a 28 días la concentración de Alita, Belita, Ferrita y Aluminato presentes en el clinker utilizado para su fabricación?

Preguntas auxiliares

1. ¿Cuál es el nivel de correlación de la concentración de las variables minerales del clinker con la resistencia a la compresión a 28 días del cemento?
2. ¿Cuál es el grado de asociación o nivel de impacto que tiene cada una de las fases minerales del clinker en las propiedades físicas del cemento, específicamente en la resistencia a la compresión a 28 días?

3. ¿Cuáles son las dos fases minerales que más impacto tienen en el resultado de la resistencia a 28 días del cemento?
4. ¿Cuál es el nivel de confiabilidad de los resultados presentados para el efecto que tiene la concentración de las principales fases mineralógicas del clinker en la resistencia a la compresión a 28 días del cemento?

Delimitación del problema

Este estudio se llevó a cabo en una empresa cementera guatemalteca que cuenta con más de 100 años de experiencia en el campo de la fabricación de cemento tipo Pórtland y de diferentes denominaciones. Con el análisis de los resultados de las concentraciones de fases minerales del clinker que ingresa al molino vertical para la producción de cemento obtenidas, por medio de un análisis de difracción por rayos X de una pastilla prensada de la muestra y los resultados de la resistencia a la compresión a 28 días de cubos de cemento preparados, según las normativas ASTM C109, muestreados a partir del promedio diario del cemento que sale del molino, resultados que fueron tomados durante un periodo de aproximadamente dos meses.

La concentración de Alita, Belita, Ferrita y Aluminato del clinker reportada con valores porcentuales y fueron calculadas, a partir de la cuantificación de los datos arrojados por el difractograma del análisis de difracción de rayos X y el valor de la resistencia a la compresión a 28 días de los cubos de cemento reportado en kN (o PSI) y, con estos resultados se establecieron las correlaciones analizadas por medio de diferentes pruebas.

OBJETIVOS

General

Determinar el grado de impacto que tienen las fases mineralógicas en el clinker en la resistencia a la compresión a 28 días en el cemento tipo UGC, mediante un análisis de correlación multivariado para muestras de cemento en un período de 30 días estudiados y presentar la propuesta de un modelo que represente dicha asociación entre variables.

Específicos

1. Determinar el grado de asociación entre la concentración de las fases mineralógicas del clinker y la resistencia a la compresión a 28 días del cemento, mediante análisis no paramétricos y correlación multivariada, para identificar los mayores niveles de correlación.
2. Jerarquizar cada una de las fases minerales del clinker para identificar el efecto de cada una de ellas en la cuantificación de la resistencia a la compresión del cemento a 28 días, mediante la aplicación del principio de parsimonia para un análisis de correlación multivariado.
3. Determinar por medio de un análisis de correlación multivariado y aplicando el principio de parsimonia, cuáles son las dos fases minerales que tienen un mayor impacto en la resistencia a la compresión a 28 días del cemento de uso general para la construcción.
4. Validar el modelo propuesto que describa la asociación de variables por medio de la aplicación de pruebas de hipótesis de los supuestos estadísticos de la regresión múltiple, para determinar la confiabilidad de la ecuación estimada y así ser aplicada en la toma de decisiones.

RESUMEN DEL MARCO METODOLÓGICO

El enfoque del estudio fue del tipo cuantitativo, porque buscó una herramienta matemática que permitiera estimar la resistencia a la compresión a 28 días del cemento tipo UGC, dependiendo de la concentración de las principales fases minerales presentes en el clinker utilizado para su fabricación.

Además, tuvo un alcance correlacional porque se asociaron los resultados del análisis químico del clinker con resultados físicos del cemento, estableciendo una correlación entre ambas y por medio de diversas pruebas estadísticas y análisis de correlación multivariada, se determinó qué nivel de asociación existe entre estas dos variables estudiadas durante un período establecido, trabajando a las mismas condiciones de operación que los equipos de una planta de cemento, ubicada en Guatemala.

La investigación realizada tiene un diseño no experimental puesto que las variables estudiadas no fueron manipuladas ni modificadas por condiciones externas, si no que fueron tratadas en su forma natural y solamente se analizaron las propiedades que tuvo cada uno de los materiales, todo esto según metodologías ASTM y COGUANOR, que son las normativas que rigen la industria cementera a nivel americano, y no pueden ser modificadas las condiciones de estos métodos; además de haber sido transversal pues las propiedades estudiadas fueron puntuales tanto de clinker como de cemento, las cuales no dependieron del tiempo sino de sus propiedades intrínsecas directamente.

La norma ASTM D75 especifica un método detallado para el muestreo del clinker, pues este depende de las condiciones e infraestructura de los equipos

de producción, se definió una población de datos correspondientes a un mesaproximadamente, que fueron tomados de un muestreo automático creando una muestra significativa de promedio diario, tanto para el clinker como para el cemento. Las muestras de clinker fueron tomadas al ingreso del molino de cemento y se analizaron por medio de la técnica de difracción de rayos X para obtener las concentraciones mineralógicas. Los datos de resistencia a la compresión a 28 días de cemento se tomaron en el laboratorio físico con la metodología normada tomando un promedio de tres resistencias por muestra.

Las variables estadísticas en la investigación para la creación del modelo predictivo, se presentan a continuación.

Tabla I. **Definición de variables independientes y dependientes**

Nombre Variable	Definición teórica	Definición operativa
Alita	Concentración de silicato tricálcico presente en el clinker	[C ₃ S]
Belita	Concentración de Silicato dicálcico presente en el clinker	[C ₂ S]
Ferrita	Concentración de ferroaluminatotetracálcico presente en el clinker	[C ₄ AF]
Aluminato	Concentración de aluminato tricálcico presente en el clinker	[C ₃ A]
Resistencia a la compresión	Resistencia en PSI que presenta el cemento a 28 días de haber fraguado en condiciones controladas de humedad y temperatura.	R28d

Fuente: elaboración propia.

El estudio se llevó a cabo en diversas fases para cumplir con todos los objetivos planteados y obtener información de confianza y relevante que brindó los aportes esperados, para la optimización del proceso de fabricación del cemento de uso general para la construcción de la cementera en cuestión.

Inicialmente se revisó toda la información bibliográfica que fue de uso para la realización del estudio, desde el análisis de investigaciones previas realizadas al cemento y al clinker, tanto para el análisis de sus propiedades físicas como químicas, ya que con esto se buscaba tener un panorama más claro de lo que se deseaba hacer, y así encontrar herramientas útiles para encaminar la investigación que se realizó, porque con ello se tomó de ejemplo lo que se ha estudiado en otros lugares y qué conclusiones obtuvieron para basar el estudio en esas experiencias.

Además de analizar estudios previos, fue de suma importancia estudiar a fondo toda la teoría involucrada en el estudio a realizar para entender el porqué de muchos fenómenos que se puedan encontrar a lo largo de la investigación.

El estudio de las técnicas y pruebas estadísticas que se utilizaron también fue uno de los pilares de la investigación, porque esta información complementó el análisis de los resultados obtenidos y cómo interpretar algunos de ellos. Conociendo cada una de las pruebas y entendiendo el objetivo para el que fueron utilizados, se fundamentó el porqué de cada una de las conclusiones dadas, ya que se explicó qué quiere decir cada uno de los resultados dados y cómo obtenerlos, también se estableció qué alcance llegó a tener cada una de las pruebas en la investigación.

Luego de contar con los fundamentos teóricos se procedió a la recolección de los datos necesarios para la investigación, los cuales fueron tomados en un período de aproximadamente dos meses, debido a que por la naturaleza de las

variables la resistencia a 28 días fueron obtenidas hasta en el segundo mes de análisis, luego de haber analizado las muestras de clinker.

Las muestras de clinker tomaron a la entrada del molino de cemento, recolectadas con una frecuencia de cuatro horas y con todas las muestras obtenidas del día se creó una única muestra de promedio diario para realizar el análisis químico. Dichas muestras fueron preparadas, según normativas ASTM para la creación de pastillas prensadas para realizar el análisis mineralógico por medio de un difractómetro de rayos X con el que se cuantificó de manera porcentual la concentración de los minerales: alita, belita, aluminato y ferrita, a través de la interpretación del difractograma obtenido.

Luego, para las muestras de cemento se procedió con la misma metodología para la creación del promedio diario de las muestras tomadas a la salida del molino vertical de cemento, con dichas muestras fueron preparados los cubos de concreto, según la metodología explicada en normas ASTM, para que fueran analizados 28 días después de su preparación y determinar el valor de la resistencia a la compresión dado en PSI, valores que se obtuvieron en el siguiente mes de análisis.

Una vez creada la base de datos, se procedió a aplicar un análisis de correlación multivariada obteniendo un modelo de 4 variables independientes; sin embargo, el modelo se simplificó con la aplicación del principio de parsimonia para disminuir la cantidad de variables independientes, según su importancia e impacto en la resistencia a la compresión a 28 días, datos que fueron recolectados y analizados en Excel, software de Microsoft Office.

Con la correlación múltiple ya simplificada, se procedió a aplicar las pruebas estadísticas necesarias para la validación del modelo propuesto; pruebas de independencia, normalidad, homocedasticidad y multicolinealidad.

Con diferentes pruebas estadísticas aplicadas a cada una de las series se pudo contar con un panorama más claro, para interpretar cada uno de los resultados obtenidos y establecer el porqué del comportamiento de las variables influyentes en la resistencia a la compresión a 28 días del cemento, y con ello proponer un modelo matemático capaz de predecir cuál será el comportamiento del cemento luego de 28 días de haber fraguado.

A cada una de las regresiones lineales individuales creadas, se aplicó una prueba de hipótesis al coeficiente de regresión, con el fin de evidenciar si el modelo propuesto cuenta con una confianza válida, según la confianza establecida para el estudio, y que fuera tomado como un modelo predictor del comportamiento de una variable física, a partir de los valores de la concentración de fases minerales encontradas por medio de difracción por rayos X de la materia prima para la producción de cemento de uso general para la construcción.

Toda la información recolectada para la realización del presente estudio, se analizó por diferentes técnicas para cumplir los objetivos propuestos y brindar una interpretación adecuada, así como contar con conclusiones que fueran realmente determinantes para brindar aportes significativos al proceso de producción de cemento de uso general para la construcción.

Las pruebas de independencia, normalidad y multicolinealidad, se aplicaron con un 95 por ciento de confianza en todos los casos que lo requirieron, usando una distribución normal Z, T-Student o F de Fisher.

INTRODUCCIÓN

El objetivo principal de la investigación es la determinación de la relación que existe entre propiedades químicas del clinker con propiedades físicas y mecánicas del cemento tipo UGC, que es fabricado con él para lo cual se presentan cuatro capítulos principales.

El primer capítulo presenta el análisis de estudios previos realizados a estas variables del clinker y cemento, ya que a raíz de la problemática que llega a existir cuando no se conoce con certeza cómo puede abordarse una investigación es importante basarse en investigaciones previas. Con esta base se estableció que, a partir de 38 muestras de clinker y 38 muestras de cemento, es posible determinar de qué manera y en qué grado, la composición mineralógica puede influir sobre el valor final de la resistencia a la compresión a 28 días del cemento UGC.

Además de contar con el respaldo de investigaciones previas, el capítulo 2 de este estudio despliega detalladamente la base literaria que contiene la información necesaria para interpretar los resultados obtenidos, indicando las fuentes de las cuales fue recolectada dicha información que luego es aplicada a los siguientes capítulos.

Los capítulos 3 y 4 presentan los resultados obtenidos en el estudio y la interpretación de cada uno de ellos respectivamente. Estos capítulos se convierten en el centro de la investigación, ya que en ellos se pretende resolver los cuestionamientos planteados en los objetivos y además explicar el por qué del comportamiento de lo calculado y presentado.

1. MARCO REFERENCIAL

1.1. Estudios previos

La resistencia a la compresión a 28 días en el cemento y cómo ésta se comporta con diferentes condiciones ha sido un tema de estudio que ha generado diversas investigaciones alrededor del mundo, debido a que se trata de una variable física crítica del cemento para presentarlo al mercado asegurando la calidad, tanto del cemento como tal, como de las materias primas y recursos utilizados para su fabricación.

Estudios de tesis, artículos científicos en revistas especializadas, así como investigaciones de instituciones dedicadas al cemento y la construcción, comprenden la mayor parte de los antecedentes sobre el estudio de esta propiedad y cómo se ve influenciada por diferentes factores. Tanto en Europa como en América se han preocupado por la optimización del proceso de producción de cemento, pero velando siempre por el cumplimiento de esta variable, según las normas internacionales estandarizadas.

España, Argentina y Colombia, han sido solo algunos de los países que han llevado a cabo diversos estudios sobre el cemento y de los cuales se han tomado referencias para la mejor realización de la presente investigación.

Vilanova Fernández en Madrid (2009), en su tesis doctoral titulada: “Influencia de la dosificación y empleo de diferentes tipos de cemento y adiciones en las propiedades mecánicas del hormigón autocompactante”, hace alusión a las variaciones que presenta la resistencia a la compresión y otras

variables físicas del concreto, según el tipo de cemento que se trabaje y cómo son adicionados los aditivos y componentes principales para la creación de concreto. Inicia el estudio presentando un análisis estadístico descriptivo de los datos obtenidos, con el fin de establecer la confiabilidad de los datos trabajados y con ellos continuar con el análisis posterior, el cual se enfatiza en la creación de correlaciones entre diferentes variables físicas con el tipo de cemento utilizado y la dosificación trabajada en el concreto. Para cada una de estas correlaciones creadas, se estimó una regresión lineal o cuadrática, según fuera el caso, mostrando el coeficiente de correlación como indicador del ajuste de los datos al modelo matemático planteado, hace referencia a que en normas se establece o se espera el comportamiento que tienen las gráficas.

Para cada una de las regresiones realizadas en dicha investigación se calculó el coeficiente de determinación R^2 para el modelo creado y se presentó como evidencia del ajuste de la correlación de los datos a la ecuación resultante. Esto brinda gran aporte al determinar que es posible generar correlaciones y regresiones de la resistencia a 28 días del cemento con otras variables no físicas; sin embargo, fue necesario estimar el grado de confiabilidad del modelo propuesto, para contar con la seguridad que la predicción de la ecuación es apegada a la realidad de los datos.

En junio de 2015, Brunatti, del Instituto de Cemento Pórtland Argentino, realizó un análisis estadístico sobre las resistencias a compresión del hormigón, comúnmente conocido como concreto, en América. Dicha investigación realizó diferentes análisis y pruebas estadísticas, con el fin de comprender cuál era el comportamiento resultante de la resistencia a la compresión, que por la naturaleza de sus resultados se espera una distribución normal para esta propiedad física del cemento. Los análisis de compresión fueron desarrollados en el departamento de control de calidad mostrando,

además del comportamiento de los resultados de la prueba de compresión por medio de la prueba de CUSUM, el momento en que esta variable sufrió un cambio significativo en su comportamiento. Una prueba de Grubb permitió detectar los datos atípicos de la distribución para así descartarlos de la serie de datos y aceptar los valores que presentaría el modelo final.

Esta investigación trabajó con tres cartas de control, basado en los objetivos del aseguramiento de la calidad de la producción que se tenían planteados y con estas herramientas de control fue posible determinar qué valores hicieron que el proceso perdiera estabilidad y dejara de cumplir con las especificaciones del producto, lo que ayudó a tomar acciones correctivas directas sobre las desviaciones detectadas. Las cartas de control son aplicadas a los datos generados por el proceso estudiado, y con estos resultados obtenidos se facilitará la toma de decisiones para la creación de metodologías, para la evaluación de la calidad del cemento; además de contar con un modelo matemático predictivo con un alto nivel de confiabilidad que podrá mostrar el comportamiento de la estabilidad y cumplimiento del proceso estudiado.

Burgos y Correa (2007) en su estudio titulado: “Diseño e implementación de una metodología estadística para ensayos de producto terminado en Cementos El Cairo, S.A.”, establecen una serie de pasos lógicos que van desde el planteamiento del problema hasta el ajuste y repetición del ciclo de análisis, pasando por la determinación del diseño experimental y la elección del tamaño de la muestra, proponiendo un método para el análisis estadístico de cualquier industria que se dedique a la fabricación de cemento, para facilitar la interpretación de la repetibilidad y reproducibilidad de sus resultados.

Investigaciones como esta aportan información para definir el inicio de un estudio, con el fin de aplicar las herramientas del análisis estadístico a cualquier

proceso o fenómeno de interés, como la correlación de una variable física contra una química, para el caso de esta investigación, evita que el producto que no cumpla con las características mínimas de calidad pueda ser despachado, y que por tratarse del mismo tipo de material, puede hacer referencia a una parte de esta investigación para realizar un análisis de la repetibilidad del proceso.

Esta investigación deja abierta la opción para futuros investigadores, el tipo de análisis a realizar que debe ser establecido en la sección de “Selección del diseño experimental” que se utilizará para la investigación, y qué pruebas validarán la confianza de la metodología aplicada a posteriores estudios.

En 2007, en Medellín, Colombia, Restrepo, Tobón y Restrepo, realizaron una investigación para determinar la influencia que presentaba la adición de diferentes minerales en el clinker utilizado para la fabricación de cemento tipo Pórtland. El análisis de difracción por rayos x fue la técnica utilizada en este estudio para caracterizar la mineralogía del clinker con los aditivos correspondientes del cemento, dicha técnica es la misma utilizada en esta investigación, debido al alto grado de confianza que genera, según lo descrito en el artículo titulado: “*Efectos de algunas adiciones minerales colombianas en la fabricación del clinker para cemento Pórtland I*”; la técnica XRD brindó los resultados necesarios y confiables para determinar cuáles eran los efectos de las diferentes adiciones de minerales, además fue utilizada como técnica de apoyo en la determinación de la temperatura de formación de minerales, junto con la aplicación de otras técnicas analíticas.

Este tipo de análisis por rayos X ha tomado gran relevancia, a través de los años, ya que es usado como la aplicación preferida en diferentes estudios de investigación de temas de cementos, tal es el caso de Henao Macías López,

Martínez & Ríos Reyes (2011) y de Castañón, García, Guerrero y Gómez (2012), quienes aplicaron la misma técnica analítica para determinar las fases minerales presentes en el clinker en su artículo: “*Estudio de las fases mineralógicas del clinker en una cementera española, utilizando el método de Rietveld*”. Las dos investigaciones dan una importancia considerable al estudio de la composición química del clinker para producción de cemento, debido a su importante efecto sobre la resistencia a la compresión del concreto.

Los resultados y las conclusiones mostradas por estos dos estudios que utilizaron la difracción por rayos X, afianzan una confiabilidad alta para los resultados que se pretende obtener en el presente estudio, ya que esta técnica ha sido utilizada para diferentes propósitos por diferentes investigadores la convierte en una herramienta clave para estudios de análisis químico del cemento. Esta investigación contó con la certeza que la técnica analítica a utilizar es moderna, actualizada y además que es utilizada frecuentemente por diferentes instituciones para la determinación mineralógica de materias primas en actividades mineras.

Los modelos matemáticos predictores de las propiedades físicas del cemento es un tema de gran interés para muchas cementeras, principalmente para los departamentos de control de la calidad y producción, porque de estas propiedades depende gran parte de las ventas y movimiento del producto y si puede predecirse de cierto modo cómo se comportarán, resultará en ahorros significativos en su fabricación.

Maury, Molinares y Sanjuán (2007), en su estudio titulado: “*Desarrollo de un modelo computacional para predecir la composición de la pasta de cemento durante el proceso de fraguado*”, recolectaron durante siete meses en la Estación Meteorológica de la Universidad del Norte en Barranquilla, Colombia,

los datos suficientes para presentar un modelo matemático predictivo del tiempo de fraguado del cemento, variable crítica a monitorear en cualquier cementera. El uso de un software adecuado, como lo fue Visual Basic, fue determinante para la creación del modelo presentado para el cual concluyen “El modelo permite predecir satisfactoriamente el estado de las diferentes fases del agua dentro de la mezcla considerando el fenómeno de evaporación de agua en la superficie del concreto en climas cálidos, y el proceso de hidratación durante el fraguado de la pasta”, (Maury, Molinares, & Sanjuán, 2007) esto da la pauta para aplicarse software que cumpla con las condiciones necesarias para la presentación de resultados, pues la tecnología es una herramienta que debe ser explotada para la mejora de las investigaciones, tal y como se aplica en esta investigación.

López y Tobón (2006) en su artículo: “Replanteamiento de las ecuaciones de Bogue en el cálculo mineralógico de clinker para una cementera colombiana” hacen una comparación entre los datos obtenidos por dos hornos diferentes que producen clinker y posteriormente hacen el cálculo teórico de la composición mineralógica que se obtiene de las ecuaciones de Bogue y los resultados reales calculados por petrografía por microscopia. Se hizo ajustes a las ecuaciones de Bogue para los resultados de cada horno trabajado, porque dichas ecuaciones asumen total pureza de las muestras, sin embargo en cada cementera sin importar su ubicación, la pureza de sus materias primas depende de diferentes condiciones, que pueden ser tanto ambientales como de la propia composición química de sus materias primas, por esta razón es que se hacen estos ajustes a las ecuaciones originales.

Las nuevas ecuaciones son obtenidas, a través de una correlación entre las variables en estudio para ajustarlas a un modelo óptimo, presentando una tabla con los coeficientes de correlación para cada caso. Sin embargo, al igual que el estudio realizado por Burgos y Correa (2007), no se presenta un análisis

practicado a las correlaciones obtenidas para determinar la confiabilidad del modelo propuesto, razón por la cual estos modelos no son totalmente aceptados como predictivos de la composición mineralógica del clinker, porque no se evidencia de manera cuantitativa un indicador para ser considerarlos como una herramienta a utilizar.

Por lo anterior, en este estudio se planteó como uno de los objetivos, evidenciar de manera cuantitativa, con la aplicación de las pruebas de hipótesis estadísticas a los datos y coeficientes de correlación calculados, el nivel de confianza de los resultados para establecer si las ecuaciones predictoras de la resistencia a la compresión a 28 días del cemento, a partir de la composición mineral del clinker, pueden ser aplicadas con la seguridad que se obtienen resultados con una aproximación significativa a la real y así tomar las mejores decisiones para ahorrar recursos en la producción de cemento, evitando la creación de producto que no cumpla con las especificaciones de calidad mínima requerida para su distribución.

Cada uno de los antecedentes referenciados han aplicado sus respectivas metodologías estudiando datos que se han recolectado a través del tiempo en plantas cementeras que cuentan con la suficiente solidez para brindar datos confiables, que fueron tomados de períodos que abarcan varias semanas para contar con resultados de mayor confiabilidad. En esta investigación se analizaron datos tomados por un período de 30 días, aproximadamente 4 semanas de producción, las cuales por la naturaleza de los resultados del análisis físico del cemento se obtienen cuatro semanas después de los datos químicos del clinker, la toma total de datos duró 8 semanas aproximadamente.

Realizar una síntesis de los aportes e información extraída de los antecedentes leídos y aplicarlos en este estudio le da un valor agregado y lo convierte en una investigación confiable, además de la literatura revisada,

cuenta con el respaldo de haber sido iniciada con ideas que investigadores previos plantearon en sus respectivos estudios del cemento realizados en diferentes países, lo que conlleva a estandarizar y asegurar que cada una de las técnicas aplicadas en la presente investigación, son las aplicadas internacionalmente y que cada resultado y dato mostrado corresponde a la correcta aplicación de metodologías normadas para la industria cementera.

2. MARCO TEÓRICO

2.1. Clinker y cemento

2.1.1. Proceso de producción del cemento

La fabricación de cemento es un proceso del cambio químico de la piedra caliza, la cual atraviesa un proceso de clinkerización para dar paso al clinker, materia prima principal para la producción del cemento, por medio de adición de otros materiales dan paso a la formación del cemento.

Según Holcim (2015), principal cadena mundial de cementeras, el proceso para la creación del cemento consta de seis diferentes etapas: obtención de materias primas, preparación de materias primas, molienda de crudo, cocción en horno rotativo, molienda de clinker y aditivos y, almacenamiento y expedición.

Con dichas etapas se completa todo el proceso de fabricación de cemento, sin embargo, implícito en dicho procedimiento, se cuenta con un monitoreo detallado para cada una de las etapas y subprocesos de la producción de cemento, el control de calidad, con el fin de respaldar que el producto que se está llevando al mercado cumple con todos los parámetros establecidos por la ASTM, quien rige las condiciones mínimas que debe cumplir el cemento; es decir, que las seis etapas principales van ligadas estrechamente con el exhaustivo control de la calidad de sus parámetros tanto químicos como físicos.

El proceso de producción de cemento se representa gráficamente en la siguiente figura.

Figura 1. **Proceso de fabricación de cemento**



Fuente: elaboración propia.

2.1.2. **Clinker**

Se conoce como clinker al material resultante del proceso de calcinación y sinterización de la harina cruda (mezcla de piedra caliza con otros minerales naturales), para llevar a cabo esta reacción se debe alcanzar temperaturas de más de 900°C para la calcinación de la materia prima y con ello obtener los óxidos necesarios en la siguiente etapa, óxido de calcio y óxido de silicio, además de otros materiales como hierro y aluminio, los cuales representan las principales fases del clinker. Una vez obtenidos los óxidos por medio de la calcinación, estos son llevados a una temperatura de 1400°C, a la cual se consigue el proceso de sinterización en el que los óxidos son mezclados para formar los minerales de alita, belita, ferrita y aluminato, que son los responsables de aportar sus principales propiedades al clinker. (American Society of Testing Materials, 2011)

2.1.3. Fases minerales del clinker

2.1.3.1. Alita (C₃S)

También conocida como silicato tricálcico, modificada en composición y estructura cristalina por la incorporación de iones ajenos, es la principal fase mineral del clinker utilizado para la creación de cemento, se encuentra presente aproximadamente entre el 30 y el 70 por ciento en masa.(American Society of Testing Materials, 2011).

2.1.3.2. Belita(C₂S)

El silicato dicálcico es la fase mineral presente entre un 15 y un 45 por ciento en el clinker de uso para la fabricación de cemento tipo Pórtland y regularmente está compuesto por su polimorfo β , el cual es monoclinico, sin embargo, otros polimorfos de este compuesto pueden llegar a presentarse en esta fase.(American Society of Testing Materials, 2011).

2.1.3.3. Aluminato (C₃A)

El aluminato tricálcico es una de las fases minerales que se presentan en el clinker para cemento Pórtland, a pesar que su presencia no supera el 15 por ciento de la masa del clinker, la importancia que tiene esta fase para la clinkerización de la harina cruda lo vuelve un compuesto principal.(American Society of Testing Materials, 2011).

2.1.3.4. Ferrita (C₄AF)

La aluminoferritatetracálcica se trata de otra de las fases principales que se encuentran presentes en el clinker, al igual que el aluminato, no supera el 15 por ciento de porcentaje en masa, su combinación con el aluminato para formar la fase líquida utilizada como medio para la clinkerización y creación de alita y belita, convierten a esta fase en determinante para la fabricación de cemento Pórtland.(American Society of Testing Materials, 2011).

2.1.3.5. Difracción de rayos X

La difracción de rayos X es una técnica de análisis de la estructura cristalina de los materiales y sus diferentes compuestos, a partir de la dispersión de los rayos X, dependiendo de la configuración de los electrones dentro de una estructura cristalina específica para cada fase existente, por medio de la detección de estas intensidades y ángulo en el que son detectadas es posible cualificar y cuantificar cada una de las fases dentro de algún material estudiado.(American Society of Testing Materials, 2011).

2.1.4. Cemento

El cemento es el producto principal utilizado en la construcción, producto de la molienda del clinker con otros materiales como yeso, puzolana, caliza, entre otros que son adicionados en diferentes proporciones dependiendo de las condiciones del clinker o del tipo de cemento que se desee producir. (Holcim, 2015)

2.1.4.1. Resistencia a la compresión

La resistencia a la compresión es una prueba física que se le practica al concreto, uso principal para el cemento tipo Pórtland, por lo tanto, esta prueba es crítica para el cemento, ya que analiza cuál es comportamiento que tiene este material en la resistencia que presenta el concreto.

Se encuentra normado internacionalmente el procedimiento a seguir para llevar a cabo este análisis, así como se encuentran estandarizados los tiempos a los cuales se debe analizar esta propiedad. La normativa ASTM expresa que esta prueba debe practicarse a cubos estandarizados en medidas y preparación, 24 horas después de fraguado el concreto, así como a 3, 7 y 28 días, en los cuales se pretende evaluar el comportamiento del cemento dentro del concreto, según el tipo de cemento fabricado, cierto tiempo de evaluación será el principalmente monitoreado. Para el caso del cemento tipo GU, la resistencia a la compresión a 28 días indica cómo se comporta el cemento ya aplicado en la construcción. (Holcim, 2015).

2.1.4.2. Parámetros de calidad

El cemento en Guatemala, como en el resto de América, se encuentra regulado por una institución internacional, con el fin de que se cumpla con ciertos parámetros mínimos para que pueda ser despachado y llevado al mercado, con el fin de garantizar el buen funcionamiento del producto y evitar que se presenten problemas graves que incluso puedan involucrar vidas humanas. La norma ASTM que cuenta con todos los límites que debe cumplir el cemento, en sus variables, tanto químicas como físicas, es la ASTM 1157 y en ella se detallan los límites que cada parámetro tiene para ser despachado y puesto a disposición de los clientes.

2.2. Fundamentos estadísticos

2.2.1. Regresión múltiple

En muchas ocasiones una sola variable independiente es insuficiente para predecir el comportamiento de una variable dependiente, por lo que necesita del aporte de otras variables independientes y cuando se crea un modelo, a partir de un conjunto de variables para poder estimar el valor de una variable respuesta es cuando se aplica la regresión múltiple, que es el resultado de diversas variables y su efecto en una sola variable dependiente.

Todos los modelos de regresión múltiple creados a partir de una base de datos, pueden representarse por medio de la siguiente ecuación general:

$$Y_i = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + \epsilon \quad [Ec. 1]$$

En dicha ecuación, Y_i viene dado por una cantidad de K variables independientes que a su vez están acompañadas de un coeficiente B, el cual indica el peso o el grado de influencia que cada una de las variables independientes tiene sobre el valor Y, además de contar con un valor B_0 y ϵ en el cual el primero representa una variable aleatoria y el segundo representa los errores, ambas variables son utilizadas para representar lo que los coeficientes de las variables independientes no son capaces de explicar dentro del modelo. (Noste, 2013)(Levine & Rubin, 2004)

2.2.2. Supuestos estadísticos de la regresión múltiple

Los supuestos de la regresión lineal múltiple son aquellas pruebas que permiten determinar la validez del modelo creado, analizando desde sus

coeficientes hasta sus residuos, con el cumplimiento de estos supuestos es posible determinar si el modelo es predictor de la variable dependiente o no. (Fuentes, 2011)

2.2.2.1. Independencia

Este supuesto se refiere al comportamiento que deben presentar los residuos de un modelo de regresión lineal múltiple, los cuales deben presentar total independencia entre ellos, es decir que, no sean producto de una tendencia dentro del modelo. (Lind, 2008) (Spiegel, 2009)

La prueba de independencia sirve para establecerla existencia de alguna colinealidad entre las variables del modelo estudiado.

La independencia hace referencia a que todas las variables que se pretende incluir en el modelo como una variable predictora, no dependan de ninguna otra variable incluida en el modelo de regresión, porque de lo contrario, dicha variable no podrá ser utilizada.

Cuando una variable independiente tiene alguna relación con cualquier otra del modelo o depende del comportamiento de otra, se dice que dichas variables presentan colinealidad y dicho fenómeno debe ser evitado en un modelo de regresión lineal múltiple. (Palomo, 2011)

2.2.2.2. Homocedasticidad

La homocedasticidad es uno de los supuestos de los modelos de regresión lineal que se refiere a que la varianza de los errores presentados

por el modelo predictivo creado, presentan una varianza constante.(Levine & Rubin, 2004)

$$V(\epsilon_i) = \sigma^2 \quad [Ec. 2]$$

La homocedasticidad del modelo multivariado plantea la hipótesis estadística que la variación de los residuos será independiente de las variables explicativas o independientes del modelo multivariado.

Para la determinación de la homocedasticidad estadística de un modelo multivariado se graficarán los valores pronosticados del modelo contra el cuadrado de los residuos para buscar la independencia entre estas variables por medio de la gráfica. (Islas, 2010).

2.2.2.3. Normalidad

Este supuesto estadístico de la regresión lineal múltiple establece que todos los residuos de la comparación entre el valor esperado con el valor observado de una serie de datos, tienen una tendencia aproximadamente a normal, con una media igual a cero.(SPSS, 2010)(Kutner, Nachtsheim, Neter, & Li, 2005)

$$\epsilon \approx N(0, \sigma^2) \quad [Ec. 3]$$

Para determinar la normalidad de los residuos se aplica la prueba de Kolgomorov-Smirnov, en la cual se plantea como hipótesis nula, la falta de normalidad de los residuos del modelo creado cuando se estima un valor de p menor a la significancia establecida para el modelo, en la mayoría de casos siendo este valor un 0.05 proveniente del 95 por ciento de confianza.

2.2.2.4. Multicolinealidad

Este supuesto también es conocido como no-colinealidad y se refiere a que dentro de las variables independientes del modelo de regresión lineal no debe existir ningún tipo de correlación entre ellas, es decir, una variable independiente no dependa de otra u otras variables independientes. (Kutner, Nachtsheim, Neter, & Li, 2005)(Levine & Rubin, 2004)

La multicolinealidad puede ser de dos tipos:

- a) Multicolinealidad perfecta: una de las variables independientes tiene una regresión lineal con una o más variables independientes.
- b) Multicolinealidad parcial: este tipo de colinealidad resulta cuando existe una alta correlación entre las variables independientes del modelo de regresión.

La multicolinealidad se observa cuando las variables independientes del modelo creado se expresan como una combinación lineal:

$$0 = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n \quad [Ec. 4]$$

Una de las formas más sencillas para determinar la colinealidad entre variables independientes es con la determinación del coeficiente de determinación R^2 para cada variable comparada con las demás. Al ser detectada cualquier dependencia entre las variables independientes, éstas deben ser eliminadas del modelo creado, porque de no ser así esta colinealidad afectará en incrementar la varianza de los coeficientes estimadores de la ecuación. (Kutner, Nachtsheim, Neter, & Li, 2005)

2.2.2.5. Prueba de hipótesis

En el amplio campo de la investigación, en muchas ocasiones se vuelve vital contar con un criterio para aceptar o rechazar algún supuesto planteado

sobre alguno de los parámetros estadísticos, este supuesto recibe el nombre de “hipótesis estadística” que se ha convertido en una de las herramientas de mayor peso en la inferencia estadística, porque permite tomar una posición fundamentada al momento de concluir sobre las características de un modelo. A través de las llamadas “Pruebas de hipótesis”.(Levine & Rubin, 2004)

En estadística se aplican diferentes tipos de pruebas de hipótesis para estimar las características de una distribución, hipótesis nula y la hipótesis alternativa. La primera es la que regularmente se plantea para alguno de los parámetros que se desee evaluar, mientras que la hipótesis alternativa es aquella que contradice lo que la hipótesis nula plantea, es decir, que la hipótesis alternativa pasará a ser aceptada al momento de rechazar la hipótesis nula que se planteó al inicio de la prueba.

Para la determinación de las conclusiones de cada una de las hipótesis propuestas en los supuestos, las pruebas de hipótesis son fundamentadas en distribuciones teóricas en las cuales se comparan dos valores, uno calculado con los valores reales de la distribución creada y otro valor teórico buscado en las diferentes tablas, según sea la prueba que se esté aplicando.

Las principales distribuciones utilizadas para las pruebas de hipótesis son las siguientes: T-Student y distribución normal (Z)

- **T-Student**

La prueba T-Student fue creada en 1899, por William Sealey Gosset, químico inglés que se dedicaba al control de calidad, quien publicó sus

hallazgos en el tema de estadística anónimamente con el nombre de “Student”.

Esta prueba es utilizada para distribuciones que atienden a la normalidad y en la cual se pueden contrastar hipótesis sobre medias, también puede ser utilizada para distribuciones que no son precisamente normales cuando se trabaja con grandes poblaciones, sin embargo no es recomendable usar esta prueba para esos casos. Esta distribución está determinada a ser usada para muestras pequeñas ($n < 30$).

Para la distribución de T-Student es importante establecer que la curva en la cual se distribuyen los datos depende de un factor importante, los grados de libertad, los cuales corresponden a un número entero positivo que representa la cantidad de variables que están libremente determinadas bajo el hecho que la suma de las diferencias entre las observaciones y el valor medio debe dar como resultado siempre cero.

La curva de la distribución T-Student varía, según la cantidad de grados de libertad que presente la serie de datos, pero ésta no se aleja de la similitud con una campana Gaussiana, también cuenta con el centro o valor medio cero y una dispersión simétrica hacia los lados; sin embargo, esta dispersión de los datos es mayor con una cantidad de grados de libertad bajo. A medida que los grados de libertad de la distribución aumenten, la campana de distribución T-Student se asemejará más a una distribución normal estándar.

Para la evaluación de pruebas de hipótesis con la distribución T-Student se hace una comparación entre dos valores de T dados por las especificaciones de un valor alfa (α) y los grados de libertad calculados para

la serie de datos, valor que es determinado por tablas ya creadas para diferentes niveles de confianza y grados de libertad que trabaje la distribución analizada, con un valor de T calculado a partir de la siguiente ecuación:

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad [Ec. 5]$$

Dependiendo de los criterios que la prueba de T-Student utilizada plantee, la hipótesis nula propuesta puede ser aceptada o rechazada, según sean los valores para la T teórica y el valor de la T calculada. (Reyes)(Noste, 2013)

- **Distribución normal**

La distribución fue estudiada por el matemático Gauss, la cual se trata de una variable aleatoria continua que puede tomar cualquier valor, siempre y cuando éste se encuentre dentro de los números reales.

Esta distribución también es llamada distribución de Gauss o gaussiana y se basa en dos parámetros, la media y la desviación estándar para las cuales se asumen valores de 1 y 0 respectivamente, esto gráficamente puede representarse por una campana simétrica (Campana de Gauss) para la función densidad de los datos de la serie.

La distribución normal, para el planteamiento de las pruebas de hipótesis, hace una comparación entre dos parámetros Z, los cuales son trazados en el intervalo de la campana y según la naturaleza de la hipótesis planteada ésta puede ser rechazada o aceptada.

Los valores de Z utilizados son calculados a partir del valor de la probabilidad analizada y buscada en la tabla ya creada para la distribución normal, que va de valores de $-\infty$ a $+\infty$ y el valor calculado de Z se determina de la siguiente manera:

$$Z = \frac{X - \mu}{\sigma} \quad [Ec. 6]$$

Con estos dos valores de Z se puede aceptar o rechazar la hipótesis nula planteada, dependiendo si el valor de Z que fue calculado queda dentro de la zona de rechazo establecido por el valor de Z dado por la tabla de distribución normal. (Kutner, Nachtsheim, Neter, & Li, 2005)(Noste, 2013)

2.2.3. Principio de parsimonia

El principio de parsimonia también es conocido como de “Navaja de Ockham” nombre, debido a su creador, un fraile, filósofo y lógico franciscano del siglo XIV, Guillermo Ockham, quien formuló la teoría que establece que si para la explicación de un fenómeno se cuenta con varias hipótesis propuestas, lo más lógico sería escoger que quien lo explica mejor es la más sencilla de todas. (Barón, 2004).

En estadística este principio es usado para simplificar modelos de regresión lineal múltiple, a manera que basados en la colinealidad de las variables independientes que forman el modelo pueda reducirse el número de ellas y crear un modelo matemático predictivo más simple, más fácil de validar y que de igual manera sea representativo de lo que se desea pronosticar. (Morales, 2012).

3. PRESENTACIÓN DE RESULTADOS

11.1. Determinación del grado de asociación entre las variables químicas del clinker con las variables físicas del cemento UGC

Se aplicó un análisis de la varianza para los 38 datos utilizados para la creación del modelo, en el cual se presenta cuál es el impacto que cada una de las variables presenta en la resistencia a la compresión a 28 días para el concreto creado con el cemento producido.

Tabla II. **Análisis de la varianza del modelo global propuesto**

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	4	5839916.899	1459979.225	12.006	3.943E-06
Residuos	33	4012801.969	121600.0597		
Total	37	9852718.868			

Fuente: elaboración propia.

Tabla III. **Análisis de la varianza por variable para el modelo propuesto**

	Coefficientes	Error típico	Estadístico t
Intercepción	252.709	1284.851	0.197
Alita	57.514	29.486	1.951
Belita	98.324	37.790	2.602
Ferrita	307.924	166.911	1.845
Aluminato	-234.756	170.953	-1.373

Fuente: elaboración propia.

Con los resultados obtenidos del análisis de la varianza, tanto para el modelo general como para cada una de las variables independientes de manera

individual, se procedió al análisis de los contrastes globales e individuales para verificar la aceptación o rechazo del modelo y variables incluidas.

Para ello se utilizó la prueba de hipótesis basada en el valor de r^2 donde la hipótesis nula establece que éste tiene un valor igual a cero, y la hipótesis alternativa expresa que es diferente de cero, por lo que sí existe una correlación entre las variables analizadas.

Tabla IV. **Análisis de contrastes globales del modelo propuesto**

CONTRASTE GLOBAL, $H_0: r^2 = 0$; $H_a: r^2 \neq 0$						
Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F calculado	F teórico	Conclusión
V_E	5839916.900	4	1459979.220	12.006	3.943E-06	Se rechaza H_0
V_{NE}	4012801.970	33	121600.060			
V_T	9852718.870	37	266289.699			

Fuente: elaboración propia.

Tabla V. **Análisis de contrastes individuales del modelo propuesto**

CONTRASTE INDIVIDUAL, $H_0: \beta_i = 0$; $H_a: \beta_i \neq 0$						
Parámetro	Valor	ϵ_β	T calculado	T teórico	Conclusión	Colinealidad
Alita	57.514	29.486	1.951	2.035	Se acepta H_0	Existe colinealidad
Belita	98.324	37.790	2.602	2.035	Se rechaza H_0	No existe colinealidad
Ferrita	307.924	166.911	1.845	2.035	Se acepta H_0	Existe colinealidad
Aluminato	-234.756	170.953	-1.373	2.035	Se acepta H_0	Existe colinealidad

Fuente: elaboración propia.

Según los resultados del análisis de contraste global, se confirma que sí existe alguna correlación entre la resistencia a la compresión del cemento a 28 días con las concentraciones de las fases minerales del clinker, debido a que se rechaza la hipótesis que plantea que r^2 sea igual a cero; sin embargo, los resultados de los contrastes individuales indican que únicamente la fase mineral Belita es la que tiene un impacto significativo en el resultado de dicha resistencia, dejando fuera del modelo a las demás variables, según esta metodología, presentan colinealidad.

11.2. Jerarquizar las fases minerales del clinker, según el grado de asociación que presenten con la resistencia a la compresión a 28 días del cemento UGC.

Para ello se estudiaron cada una de las fases por medio de un análisis del índice de inflación, para determinar la influencia de cada una de las fases minerales por separado. Con estos índices de inflación se logra establecer la jerarquía deseada, dado que la variable que presente el menor índice será la variable que mayor impacto tiene en la variable dependiente, en este caso la resistencia a la compresión a 28 días del cemento.

Tabla VI. **Índices de Inflación de las fases minerales del clinker para la resistencia a la compresión a 28 del cemento UGC**

	Coefficiente de correlación	Índice de inflación
Sin Alita	0.546	2.20
Sin Belita	0.509	2.04
Sin Ferrita	0.551	2.23
Sin Aluminato	0.569	2.32

Fuente: elaboración propia.

Con estos resultados del índice de inflación, se establece que la Belita es la fase mineral que mayor impacto tiene en la resistencia que presenta el cemento, hecho que ya se había presentado previamente en el análisis de varianza del modelo y se confirma con esta metodología. Los resultados de la jerarquía a partir del índice de inflación se presentan en la tabla a continuación:

Tabla VII. **Jerarquía de fases minerales, según su impacto en R28d**

Posición	Fase mineral
1ra.	Belita
2da.	Alita
3ra.	Ferrita
4ta.	Aluminato

Fuente: elaboración propia.

11.3 . Modelo de regresión propuesto con las dos fases minerales de mayor influencia en la resistencia a la compresión.

Tabla VIII. **Determinación de coeficientes para modelo predictivo de la resistencia a la compresión del cemento a 28 días**

	Coeficientes	Error típico	Estadístico t
Intercepción	-1435.517	1126.691	-1.274
Alita	92.738	18.644	4.974
Belita	155.273	30.389	5.110

Fuente: elaboración propia.

La tabla anterior muestra los valores de los coeficientes para la creación del modelo predictivo de la resistencia a la compresión a 28 días, a partir de la concentración de alita y belita, fases de mayor impacto en dicha variable, para su posterior análisis y validación por medio de los supuestos estadísticos de la

regresión lineal. Dicho modelo obtenido por medio del análisis de la varianza que se presenta a continuación.

Modelo lineal predictivo propuesto para la resistencia a la compresión a 28 del cemento tipo UGC.

$$R_{28d} = 92.738[C_3S] + 155.273[C_2S] - 1435.517$$

Tabla IX. **Análisis de contrastes globales del modelo propuesto**

CONTRASTE GLOBAL, $H_0: r^2 = 0$; $H_a: r^2 \neq 0$						
Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F calculado	F teórico	Conclusión
V_E	2	5149046.826	2574523.413	19.157	2.401E-06	Se rechaza H_0
V_{NE}	35	4703672.042	134390.6298			
V_T	37	9852718.868				

Fuente: elaboración propia.

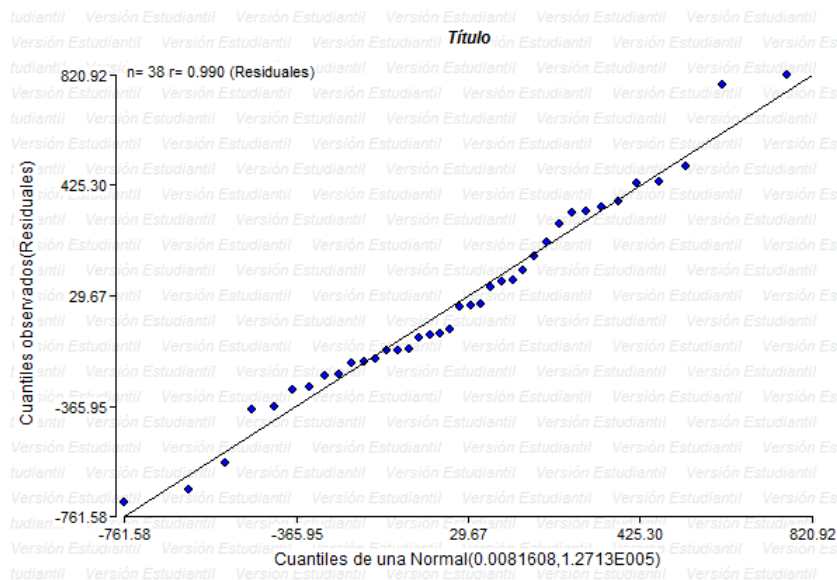
Este modelo lineal propuesto presenta un $r^2 = 0.523$, el cual es considerado significativo bajo la prueba de hipótesis del contraste global, para confirmar su validez debe ser sometido a las pruebas de hipótesis de los supuestos de independencia, homocedasticidad, normalidad y multicolinealidad para aprobar o no su aplicación en la industria cementera.

11.4 Validación del modelo predictivo propuesto, a través del análisis de los supuestos estadísticos de regresión lineal múltiple

El primer supuesto a evaluar es la normalidad de los residuos, lo cual se presenta la Figura 1, se estima normalidad si todos los puntos de la gráfica se

apegan significativamente a una recta, de lo contrario no existe normalidad en los residuos y el modelo no es validado.

Figura 2. Q-Q plot de los residuos



Fuente: elaboración propia.

El resultado de esta gráfica es una serie de puntos que se aproximan a una recta, tal y como se espera visualizar para aceptar la normalidad; sin embargo, no puede afirmarse esto, ya que es una prueba basada en el criterio del investigador, para verificar este resultado se aplicó la prueba de Kolmogorov-Smirnov a los residuos, la cual plantea en su hipótesis nula que los residuos no presentan normalidad cuando el valor del estadístico p se encuentra por debajo de la significancia trabajada, 0.05 para este caso.

Tabla X. Prueba de Kolmogorov-Smirnov para normalidad de residuos

Variable	Ajuste	Media	Varianza	n	p-valor	Conclusión
Residuos	Normal (0, 1)	0.01	127126.27	38	< 0.0001	Se rechaza Ho

Fuente: elaboración propia.

El resultado presentado en la tabla IX indica que la hipótesis nula es rechazada los residuos obtenidos de este modelo predictivo sí corresponden a una distribución normal, confirmando así lo estimado por la gráfica de Q-Q plot.

El siguiente supuesto a analizar es la multicolinealidad de las variables estudiadas para confirmar que entre las variables independientes no existe ningún tipo de relación o dependencia una de la otra, lo cual impediría aceptar dichas variables como predictoras de la resistencia a la compresión a 28 días del cemento UGC. Para analizar este supuesto, se realizó un análisis de los contrastes individuales del modelo, con el fin de encontrar estas relaciones o descartar las mismas.

La hipótesis nula de esta prueba plantea que el coeficiente calculado para la variable analizada es igual a cero, por lo que no es incluida en el modelo y para ello, el valor del estadístico T calculado debe ser menor que el valor del estadístico T obtenido a través de la tabla de T-Student, según la significancia de 0.05 y los grados de libertad obtenidos de igual forma en el análisis de la varianza del modelo.

El resultado de este análisis se presenta en la tabla XI a continuación:

Tabla XI. **Análisis de multicolinealidad de las variables**

CONTRASTE INDIVIDUAL, $H_0: \beta_i = 0$; $H_a: \beta_i \neq 0$						
Parámetro	Valor	ϵ_β	T calculado	T teórico	Conclusión	Colinealidad
Alita	92.738	18.644	4.974	2.030	Se rechaza H_0	No existe colinealidad
Belita	155.273	30.389	5.110	2.030	Se rechaza H_0	No existe colinealidad

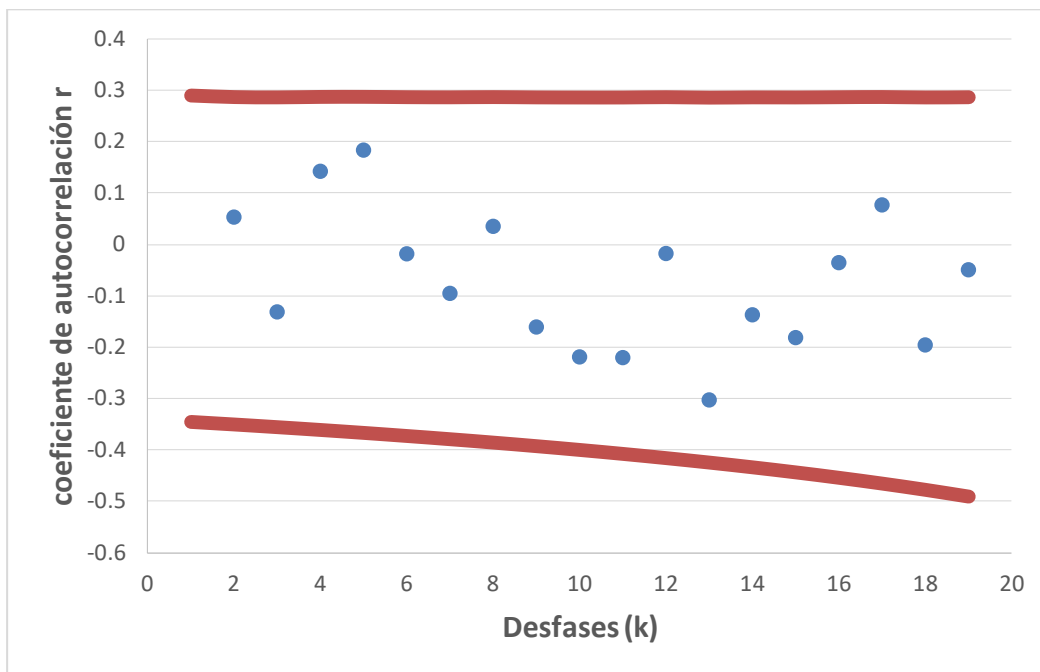
Fuente: elaboración propia.

Tal y como se confirma con este análisis, tanto la fase de la belita como la de la alita son totalmente independientes y no dependen entre sí, por lo que pueden ser incluidas en el modelo con la confianza que pueden llegar a ser predictoras para la resistencia a la compresión.

Otra de las condiciones que el modelo debe cumplir para ser validado es la independencia, para esto se aplicó la prueba de Anderson para determinar la independencia de las variables belita y alita a partir del correlograma obtenido de los valores de r y de 19 desfases creados de los 38 datos manejados para la obtención de dicho modelo.

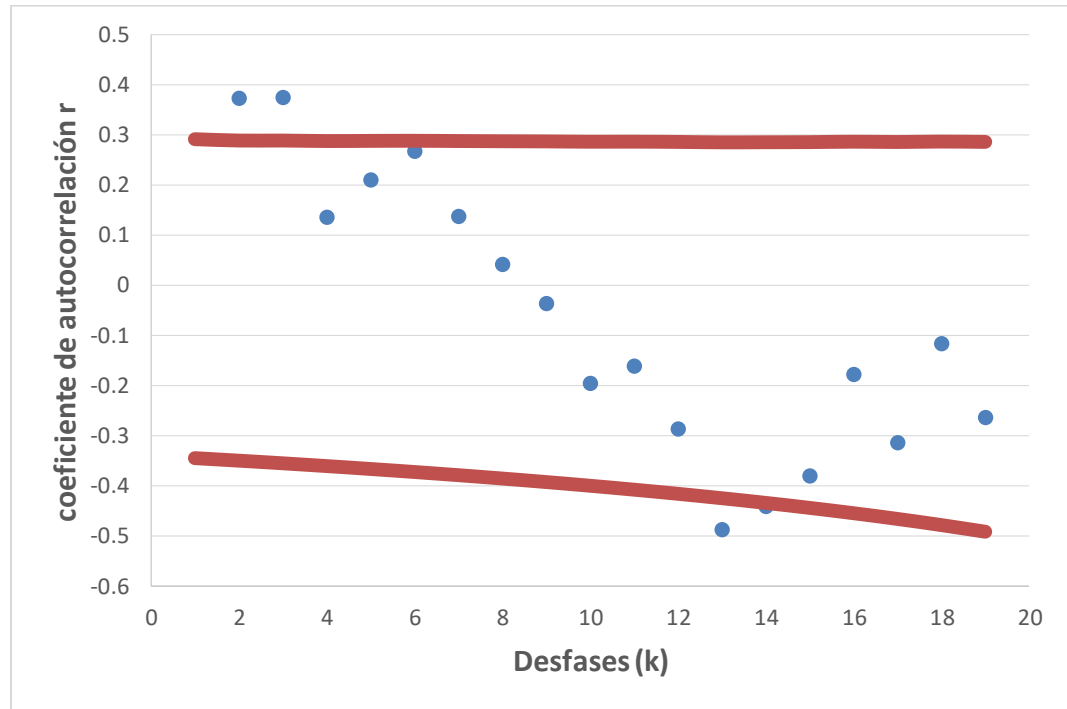
Los resultados de esta prueba se presentan en las siguientes figuras, trabajadas siempre a un grado de confianza del 95 por ciento.

Figura 3. Prueba de Anderson para la belita (C_2S)



Fuente: elaboración propia.

Figura 4. Prueba de Anderson para la alita (C_3S)



Fuente: elaboración propia.

Esta prueba es de criterio visual y dependerá del investigador si se acepta la hipótesis nula que establece que no existe independencia o si ésta es rechazada. Tanto para el caso de la belita como para la alita la mayor cantidad de puntos correspondientes a los valores r se encuentran dentro del intervalo planteado por el correlograma a 95 por ciento, por lo que se asume que ambas variables son independientes. Tal y como se ha venido analizando y comprobando en cada una de las pruebas realizadas la belita presenta mejores resultados, pues cuenta con una mayor cantidad de puntos dentro del intervalo.

La prueba de contrastes individuales realizadas para las cuatro fases minerales daba los primeros indicios que la belita era la fase que mejor explicaba el modelo y en esta prueba se reafirma esta información, pues es la que presenta una mayor independencia.

4. DISCUSIÓN DE RESULTADOS

La primera etapa de la investigación fue enfocada en determinar la relación o influencia que cada una de las fases minerales presentes en el clinker llega a tener sobre la resistencia a la compresión a 28 días del cemento tipo UGC, ya que son estas fases las que afectarán el comportamiento del cemento, debido a las propiedades que cada una de ellas aporta.

4.1. Grado de asociación entre la alita, belita, ferrita y aluminato en el clinker con la R28d del cemento.

Con el lote de 38 datos que se trabajó el primer análisis de la varianza con las cuatro variables independientes y la resistencia del cemento como variable respuesta, se logró el primer acercamiento a lo que pudo ser una ecuación parapredecir el comportamiento de dicha variable física, a partir del porcentaje presente de las fases minerales en el clinker.

Los valores y coeficientes obtenidos de este análisis permitieron un análisis de componentes globales para el modelo calculando los valores t para su comparación y evaluar las hipótesis planteadas en la cual la hipótesis nula plantea que el coeficiente de correlación entre las variables es igual a cero, por lo que el modelo no obedece en ningún momento al comportamiento real. Sin embargo, al analizar ambos valores estadísticos t, la hipótesis alternativa fue aceptada concluyendo que el modelo propuesto es aceptado, pues esta hipótesis indica que el valor del coeficiente de correlación es diferente de cero, por lo que sí existe alguna relación entre las variables propuestas.

Este análisis no puede tomarse cómo válido solamente con el contraste global estudiado, por lo que fue necesaria la evaluación de los contratos

individuales de cada una de las variables involucradas en el modelo lo que permite establecer qué variables pueden ser consideradas como aceptadas en el modelo y cuáles presentan colinealidad, por lo que no aportan información significativa para el modelo propuesto y de esa manera sean eliminadas con el fin de obtener un modelo de mayor precisión.

El resultado del análisis de los contrastes individuales trabajando con las cuatro fases minerales (alita, belita, ferrita y aluminato) produjo un resultado que llega a contradecir, lo que el análisis global presentó y es que de las cuatro fases analizadas, 3 de ellas presentan colinealidad en este modelo. Sin embargo, con este análisis es imposible determinar cuáles son las variables que tienen relación entre sí o depende una de la otra, lo que impide que la resistencia a la compresión a 28 días del cemento pueda depender al mismo tiempo de las 4 variables.

La belita es la única de las variables independientes que con este análisis rechaza la hipótesis nula, rechazando con ello que pueda tener alguna relación con las demás variables, convirtiéndola en la variable más independiente de todas, y posiblemente la que mejor pueda ajustar el modelo.

4.2. Determinación de la jerarquía de las fases minerales del clinker en la R28d del cemento UGC.

Una vez determinado que existe colinealidad entre las variables independientes, se planteó como otro de los objetivos, jerarquizar estas cuatro variables independientes para determinar cuál de todas es la que mayor impacto tiene sobre la resistencia a la compresión a 28 días del cemento y cuál de ellas es la que menos impacto genera en dicha propiedad.

Para llevar a cabo esta jerarquía la metodología utilizada fue el análisis por medio del índice de inflación obtenido para cada variable independiente, a partir de la omisión de ella en una simulación de regresión para obtener un coeficiente de determinación diferente en cada uno de los casos. El índice de inflación presenta un valor que relaciona el coeficiente de correlación obtenido para cada uno de los casos analizados con el máximo valor de correlación, uno, por lo que el análisis presenta que a mayor índice de inflación la variable anulada del análisis de regresión representa un menor impacto en la resistencia del cemento, porque el valor del coeficiente de correlación que presenta el modelo en ausencia de dicha variable, es un valor menor que el que presenta integrando la variable en el análisis de regresión.

Para este caso particular, los índices de inflación calculados así como los coeficientes de correlación están descritos en la Tabla V y tal como se explicó anteriormente, cuando se analiza el modelo sin la presencia de la concentración de belita el coeficiente de correlación es el menor de todos los casos, mientras que si se omite el aluminato del modelo el coeficiente de correlación es el más alto de los cuatro casos, provocando que su índice de inflación sea el más alto de todos, lo que lleva a asumir que esta variable es la que menor impacto genera en la resistencia a la compresión del cemento o que esta variable es la que presenta colinealidad con alguna otra de las tres variables analizadas.

Basados en los índices de inflación resultantes ya es posible crear la primera idea de cuáles variables son las que tienen mayor influencia y con un índice de inflación de 2.09 la belita es la fase mineral que aparentemente tiene un impacto de mayor significancia en la resistencia del cemento, seguida por la concentración de alita que presenta un índice de inflación de 2.20, la ferrita con 2.23 de índice de inflación queda como la tercera fase mineral de significancia para la resistencia del cemento y finalmente, el Aluminato se convierte en el

mineral, de los cuatro estudiados, que tiene un efecto menor en la resistencia a la compresión a 28 días en el cemento, que también puede explicarse que estos últimos dos minerales dependen directamente de la concentración de belita en el clinker y es por ello que al inicio el primer análisis de la varianza con todas las variables incluidas reportaba colinealidad en tres de ellas.

4.3. Determinación de las dos fases minerales de mayor influencia en la R28d del cemento

Realizada la jerarquía de la influencia que estos minerales tienen sobre la variable física del cemento, simplificar el modelo a dos variables es más sencillo y dados los resultados del índice de inflación las fases minerales a trabajar serán la belita y la alita, ya que son las que, según el análisis, tienen un efecto de mayor impacto en la variable de interés.

Una vez seleccionadas las variables que se estudiarán, se creó el modelo lineal que mejor pueda describir el comportamiento de la resistencia a la compresión a 28 días del cemento, a partir de los resultados del análisis de varianza para estas dos fases mineralógicas. Los resultados de este análisis se presentan en la tabla VII, donde se encuentran los coeficientes para cada variable, así como el intercepto para la ecuación. Tal y como se presenta en el tercer objetivo de la investigación, la selección de las dos fases de mayor influencia en la resistencia para la simplificación del modelo hace más fácil su creación y validación, además que el modelo no pudo ser construido con todas las variables debido a la colinealidad que se presentó en algunas de ellas y fueron excluidas de la ecuación de regresión.

Los resultados del análisis de contrastes individuales de las cuatro fases y el índice de inflación que logró jerarquizar cada una de ellas, confirma que tanto la ferrita como el aluminato son fases que llegan a tener relación con alguna

otra variable predictora y son las que menos influencia tienen en el resultado de la resistencia a la compresión. Químicamente se sabe que tanto la ferrita como el aluminato son fases que al momento de la clinkerización se convierten en una fase líquida que es utilizada como el medio ideal para llevar a cabo la transformación de las otras dos fases minerales estudiadas resultado acorde a lo que estadísticamente, se determinó por los análisis realizados, puesto que tanto la alita como la belita dependerán de la concentración de estas fases menores para variar su presencia dentro del clinker utilizado para el cemento.

El modelo propuesto, indicado en la sección de resultados, permite la estimación del comportamiento de la resistencia a la compresión a 28 días del cemento tipo UGC, dado en PSI, para variables de entrada de concentración porcentual de los minerales belita y alita, dichos valores se estiman con un coeficiente de correlación de 0.523, el cual, según el análisis de varianza, es aceptado pues la prueba de hipótesis planteada acepta que dicho coeficiente es diferente de cero dando lugar a la idea propuesta que la resistencia a la compresión del cemento, en parte, sí dependerá de las concentraciones minerales presentes, mas no es de lo único que depende y por eso el valor del coeficiente de correlación no es mayor.

Al trabajar únicamente con las dos variables que tienen un impacto más grande en la resistencia de cemento es imperativo analizar nuevamente las varianzas de las variables, con el fin de establecer si dichas variables pueden ser aceptadas o no, y tal como se presenta en la tabla VIII, el contraste global para el modelo bivariado propuesto rechaza la hipótesis nula, aceptando que el modelo tiene un coeficiente de correlación diferente de cero. Con este análisis se confirma lo obtenido previamente con el análisis de regresión de estas variables, confirmado de esta manera que dicho modelo sí puede predecir, de cierta manera y hasta cierto grado de significancia, el comportamiento que

tendrá la resistencia a la compresión a 28 días del cemento tipo UGC, basado en la concentración de sus fases minerales.

Entender la composición química del cemento se vuelve una base importante para confirmar o rechazar los resultados obtenidos por medio de estas pruebas, y como se ha estudiado previamente, la resistencia del cemento se le acredita principalmente a la fase de silicato tricálcico (alita) y en una proporción menor al silicato dicálcico (belita). Con esta información basada en la literatura y antecedentes, es factible confirmar que el modelo propuesto utiliza las fases minerales de mayor impacto, ya que fue corroborado con metodologías estadísticas demostradas por medio de pruebas de hipótesis; sin embargo, se debe considerar que, según las 38 muestras utilizadas, la belita es la fase que mejor describe el comportamiento de la resistencia a la compresión a 28 días y luego de ella está la alita.

4.4. Validación del modelo predictivo propuesto

Un modelo de regresión lineal, sea simple o multivariado, no puede asumirse como correcto con este simple análisis, todo modelo deberá cumplir con ciertos supuestos estadísticos que le brindan al usuario la información mínima necesaria para concretar su análisis, en este caso particular, sobre estos dos minerales que se tomaron para la predicción.

Los supuestos a evaluar para la validación de este modelo son: normalidad, multicolinealidad e independencia. Para ello, se aplicó las pruebas correspondientes para cada una de ellas, con el fin de establecer la aceptación o rechazo de las hipótesis planteadas.

El supuesto de normalidad establece que los residuos obtenidos del modelo de regresión deben tender a la normalidad y no deben obedecer ningún

patrón para ser aceptado el modelo. Para ello se realizaron dos pruebas, la primera es la gráfica Q-Q plot presentada en la figura 2 en la que los valores graficados para los residuos obtenidos, a partir del modelo creado se apegan lo suficiente a la recta definida para esta prueba, evidenciando así, la normalidad de los residuos; además de esta prueba gráfica se realizó la prueba de Kolgomorov-Smirnov basados en valores de T-Student para definir por medio una prueba de hipótesis lo ya definido por prueba anterior, donde el p-valor es menor a 0.05, utilizado como valor confianza, rechazando la hipótesis nula y aceptando la normalidad de los residuos del modelo propuesto.

La Tabla XI contiene los resultados de los contrastes individuales de las dos variables utilizadas para la creación de este modelo matemático, y es importante recalcar la diferencia que se marca en el nuevo análisis, ya que en este caso ambas concentraciones rechazan la hipótesis que establece que existe colinealidad entre las variables predictoras, con estos resultados el modelo es confiable, tanto la concentración de alita, como de belita no dependen una de la otra y las dos pueden ser usadas sin problemas para predecir la resistencia a la compresión del cemento UGC.

Finalmente, se evaluó la independencia de las variables utilizadas en el modelo, para ello el criterio utilizado fue el análisis de la prueba de Anderson para cada una de las variables utilizadas en esta aplicación.

La figura 3 representa el correlograma para la variable belita, con los 18 puntos obtenidos de los desfases trabajados se encuentran dentro de los límites que esta prueba establece, confirmando que esta variable es totalmente independiente, es decir que no depende de ningún otro factor su valor, siendo una excelente opción para ser utilizada en un modelo de regresión de este tipo.

Por otro lado, la figura 4 es la representación del correlograma de la prueba de Anderson para la variable alita, en la cual de los 18 puntos presentados un porcentaje bajo se encuentra fuera de los rangos de aceptación de la prueba, sin embargo, más del 85 por ciento de los puntos se encuentran dentro de los límites; además que los puntos fuera de esta zona no se encuentran lejanos significativamente, es una delta mínimo el que tienen para poder encontrarse en los límites de aceptación; esto indica que a pesar de que la alita no es totalmente independiente, como la belita, hay factores no controlados dentro de este proceso que hacen que esta variable tenga cierta dependencia de algunas condiciones desconocidas, pero regresando al análisis previo de multicolinealidad se asume que esta dependencia no es de ninguna de las variables utilizadas en el modelo.

Un análisis a nivel microscópico de las composición de las fases minerales dentro del clinker aportaría una mejor interpretación química y avalaría la aleatoriedad de los resultados obtenidos de manera estadística para la belita y alita, dado que, además de la cantidad presente de cada uno de estos minerales dentro de la muestra, también es importante establecer cómo está distribuido y formado a nivel de partículas.

Al llevar a cabo todo el proceso de análisis de datos históricos producidos por el departamento de control de calidad, se vuelve un factor determinante la manera en que la estadística comprueba, a partir de sus propias metodologías matemáticas, lo que según la ciencia de la química presenta en su literatura y la forma en que se aplica un análisis de regresión lineal múltiple para demostrar información dada por otra ciencia hace que la estadística se vuelva una herramienta vital que debe ser implementada y aplicada no solo a nivel de control de calidad, si no que en otras áreas y departamentos dentro de un proceso de producción.

CONCLUSIONES

1. El análisis estadístico de regresión lineal múltiple logró determinar que existe relación entre la concentración de los minerales: alita, belita, ferrita y aluminato con la resistencia a la compresión a 28 días del cemento tipo UGC; cuyo análisis de varianza propone un modelo matemático aproximado válido con solamente dos fases minerales principales (belita y alita) para predecir dicho valor, a través de la concentración porcentual de estas fases en el clinker utilizado para la fabricación de cemento.
2. Un análisis estadístico de regresión lineal multivariado evidenció que el valor de la resistencia a la compresión del cemento sí es explicado por la concentración mineral en el clinker de alita, belita, ferrita y aluminato; sin embargo, no todas las variables generan el mismo impacto en R28d.
3. La jerarquización del impacto que la concentración de las variables independientes estudiadas tiene sobre la R28d fue determinada por el valor que presentó el índice de inflación estadístico de cada una de ellas, siendo este resultado, en orden ascendente, el siguiente: aluminato, ferrita, alita y belita, entonces esta última es la que mejor describe el comportamiento de la R28d del cemento UGC.
4. La concentración de los minerales belita y alita son los que linealmente presentan un mayor ajuste significativo en el modelo de la resistencia a la compresión a 28 días del cemento tipo UGC.
5. El modelo matemático propuesto, como resultado del análisis de regresión de la base de datos estudiada es: $R28d = 92.738[C_3S] + 155.273[C_2S] -$

1435.517, donde $[C_3S]$ representa la concentración de alita y $[C_2S]$ es la concentración de belita; y una vez evaluado y aprobando cada uno de los supuestos estadísticos de regresión lineal múltiple que confirman, validan y sustentan se permite su uso para la determinación de la resistencia a la compresión a 28 días para el cemento tipo UGC siempre y cuando sea evaluado en las mismas condiciones en las que obtuvo los datos utilizados para su aproximación.

RECOMENDACIONES

1. Incluir otras fases minerales, como la cal libre, obtenidas por medio de la difracción de rayos X que inicialmente son consideradas de poco impacto en las propiedades finales del cemento para generar nuevas tendencias sobre el impacto real de dichos compuestos.
2. Jerarquizar las mismas fases minerales analizadas en este estudio, para las resistencias iniciales traería a discusión la importancia de otras fases, ya que con ello puede determinarse si existen cambios en los resultados del impacto que cada fase tiene para el comportamiento de resistencias a diferentes tiempos.
3. Analizar qué sucede con las otras dos variables y qué efecto o influencia tienen sobre la resistencia a 28 días es otro tema que resultaría importante para mejorar la calidad del producto final, así como establecer parámetros para las mismas y optimizar aún más el proceso en general.
4. Utilizar el modelo de regresión creado como guía para la determinación de la resistencia a la compresión a 28 días del cemento tipo UGC, mas no como un modelo que pueda determinar con total acierto el valor, en pulgadas por pie cuadrado, que la muestra analizada pudiera obtener porque debe considerarse que dicha propiedad depende de otras variables que no están siendo consideradas en el modelo predictivo propuesto, que generan una idea del posible comportamiento de dicha propiedad de calidad del cemento.

REFERENCIAS BIBLIOGRÁFICAS

American Society of Testing Materials. (Diciembre 2011.) *Determination of the proportions of phases in Portland cement and Portland-Cement Clinker using x ray powder Diffraction analysis*. West Coshohocken, Estados Unidos.

Acuña, E. (s.f.) *Regresión lineal múltiple* presentado en el Departamento de Matemática de la Universidad Autónoma de Puerto Rico

Azor, J. (s.f.) *Ajuste de curvas* presentado en la cátedra de Estadística II de la Universidad de Mendoza, Argentina

Azor, J. (s.f.) *Funciones Aleatorias* presentado en la cátedra de Estadística II de la Universidad de Mendoza, Argentina.

Barón, J. y Téllez, F. (diciembre, 2004) *Apuntes de Bioestadística* presentado en el tercer ciclo de la carrera en Ciencias de la Salud y Medicina, Universidad de Málaga, España.

Burgos, C. & Correa, A. (2007). Diseño e implementación de una metodología estadística para ensayos de producto terminado en cementos El Cairo, S.A. *DYNA*, 74(151).

Brunatti, C. (2015). *Análisis de Datos- control de calidad*. Instituto del Cemento Portland Argentino. Montevideo.

Castañon, A., García, S. Gómez, F. & Guerrero, A. (2012). Estudio de las fases mineralógicas del clinker en una cementera española, utilizando el método de Rietveld. *DYNA*, 79(173), 41-47.

Fuentes, S. (2011) *Regresión lineal*, presentado en la Facultad de Ciencias Económicas y Empresariales de la Universidad Autónoma de Madrid.

Henao, J., Macias, M. y Ríos, C. (2011). Análisis estructural de la analcima obtenida a partir de clinker natural (NC-ANA), mediante difracción de rayos X de muestras policristalinas y espectroscopia de resonancia magnética nuclear de estado sólido de ^{29}Si y ^{27}Al . *Revista de la Facultad de Ingeniería Universidad de Antioquia*, (61), 9-18.

Hispano Portuguesa SPSS (s.f.) SPSS Guía para el análisis de datos, en análisis de regresión lineal: el procedimiento regresión lineal (Cap. 18 pp 1-69) Madrid, España.

Holcim. (2015). *Holcim España*. Obtenido de <http://www.holcim.es/editorials/proceso-de-fabricacion.html>

Islas, F. y Vela, F. (junio, 2010) *Regresión lineal multivariada*, presentada en los Talleres de verano en métodos de investigación en población y territorio. Universidad Autónoma Metropolitana, Unidad Xochimilco, México D.F.

Kurtner, M. Nachtsheim, C., Neter, J. & Li, W. (2005) *Applied Linear Statistical Models*, (5th Ed.) New York: McGraw-Hill Irwin.

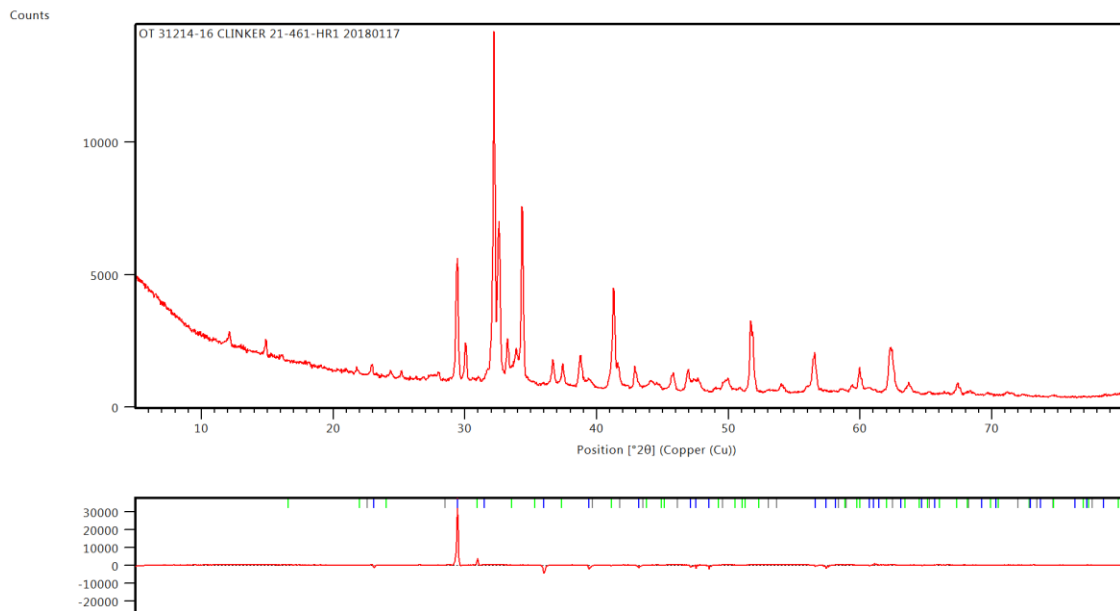
Levine y Rubin (2004), *Estadística para Administración y Economía* (7ma edición) México: Pearson Education

- Lind, D., Marchal, W. y Wathen, S. (2008) *Estadística Aplicada a los Negocios y Economía* (13 ed.) México: McGraw-Hill Interamericana.
- López, F. y Tabón, J. (2007). Replanteamiento de las ecuaciones de Bogue en el cálculo mineralógico del clinker para una cementera colombiana. *DYNA*, 74(153), 53-60.
- Maury, A., Molinares, N. & Sanjuán, R. (2007). Desarrollo de un modelo computacional para predecir la composición de la pasta de cemento durante el proceso de fraguado. *DYNA*, (22).
- Morales, A. (2012) *Estadística y probabilidades*, presentado en la Universidad Católica de la Santísima Concepción, Chile.
- Noste, M (2013) *Apuntes de regresión lineal* presentado en la Facultad de Ciencias exactas y naturales para la carrera de Especialización en estadística para Ciencias de la Salud. Buenos Aires, Argentina.
- Palomo, J. (julio, 2011). *Regresión lineal múltiple*. Trabajo presentado en la Escuela Universitaria de Arquitectura Técnica de la Universidad Politécnica de Madrid.
- Restrepo, J., Restrepo, O. & Tobón, J. (2007). Efecto de algunas adiciones minerales colombianas en la fabricación de clinker para cemento Pórtland I. *DYNA*, 74(152).
- Spiegel, M. y Stephens, L. (2009) *Estadística* (4ta ed.) México: McGraw-Hill Educación.

Vilanova Fernández, A. (2009). Influencia de la dosificación y empleo de diferentes tipos de cemento y adiciones en las propiedades mecánicas del hormigón autocompactante. (Tesis de doctorado). Universidad Politécnica de Madrid, Madrid.

APÉNDICES

Apéndice 1. Ejemplo de difractograma de clinker



Fuente: elaboración propia.