

**UNIVERSIDAD NACIONAL AUTONOMA DE
NICARAGUA -LEON
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA-MATEMATICA**



**DETERMINACION DE LOS INTERVALOS DE CONFIANZA DE LAS
VARIABLES: CENIZA, HUMEDAD Y PROTEINA EN EL PROCESO
DE CONTROL DE CALIDAD EN LA PRODUCCION DE LA HARINA
EN LA INDUSTRIA GEMINA S.A.
EN EL PERIODO AGOSTO-DICIEMBRE 2001**

Presentado por:

Bra. MARTA DIXON REYNALES.

Tutora: MSc. Ana Cristina Rostrán Molina.

**Asesora: Lic. Nidia Chui Blandino
Departamento de Control de Calidad GEMINA S.A.**

León, Enero de 2002.

190.530
C.3

EST
378.2
D621d
2002

INDICE

I.	INTRODUCCIÓN	6
II.	OBJETIVOS	8
III.	MARCO TEORICO	9
	3.1.1. - PROCESO DE ELABORACIÓN DE LA HARINA	9
	3.1.2. - Costo de Calidad	11
-	3.1.3.- Variables y Estándares del Proceso de Producción de la Harina.	13
	3.1.4. - Producto y Servicios Principales que ofrece GEMINA S. A.	13
	3.2. Técnicas Estadísticas para el Análisis de Datos	15
	3.2.1.- El intervalos de Confianza.	15
	3.2.2.- Análisis de Correlación Lineal	16
	3.2.3.- Análisis de Varianza.	16
	3.2.4.- Modelo Lineal.	16
	3.2.5.- Análisis de Regresión Lineal Múltiple	17
	3.2.6.- Definición del Modelo de Regresión Lineal Múltiple.	18
	3.2.7.- Hipótesis a Priori sobre la Distribución de los errores.	18
	3.2.8.- Las hipótesis respecto a la perturbación puede Escribirse en términos de las variables respuestas.	19
	3.2.9.- Especificación del Modelo.	20
	3.2.10.- Prueba D de Durbin-Watson.	21
IV	MATERIALES Y METODOS.	23
V	RESULTADOS Y DISCUSION.	24
	5.1. - Distribución de las variables.	24
	5.2. - Estadísticos Descriptivos.	27
	5.3. - Análisis de Regresión y Correlación.	35
	5.3.1.- Correlaciones Simples.	35
	5.3.2.- Análisis de Regresión Múltiple.	38
VI.	CONCLUSIONES.	43
VII	RECOMENDACIONES.	45

VIII.	BIBLIOGRAFIA.	46
IX.	ANEXOS.	47
	ANEXO 1. Hojas de Registro de la Empresa GEMINA S.A.	48
	ANEXO 2. Diagrama de Bloque	50
	ANEXO 3. Pruebas de Normalidad	51

DEDICATORIA.

Este trabajo se lo dedico con mucho amor y cariño.

A MI ESPOSO.

Por haberme brindado su apoyo en situaciones difíciles supo comprenderme, de alcanzar esta meta final.

A MI HIJO.

Le agradezco mucho a mi hijo el cual me dio valor a estudiar esta carrera.

AGRADECIMIENTOS.

A DIOS.

Le Agradezco Al Dios Todo poderoso de iluminarme la mente, por haberme dado la fuerza y sabiduría para culminar mis estudios.

A MIS PROFESORES.

Les agradezco mucho por haberme brindado sus mayores conocimientos.

A MI TUTORA.

A la Msc. Ana Cristina Rostrán, por darme su confianza para realizar el presente trabajo por compartir sus conocimientos Profesionales. Por haberme brindado su valioso tiempo y dedicación para el desarrollo del mismo.

Le agradezco a la Lic. Nidia Chui Blandino, Resp. De Control de Calidad en la Empresa GEMINA, Por todas sus atenciones y brindarme la información de laboratorio.

A los estudiantes de cuarto año de Estadística de clase a Julio Reyes S. Douglas Solis y Denis Neyra quienes me apoyaron con la información.

I.- INTRODUCCIÓN

Nicaragua enfrenta el reto de competir con sus productos en los mercados locales e internacionales para ello necesita establecer los parámetros de control de las variables que interviene el proceso productivo.

La industria GEMINA S.A. es integrante del sector secundario que esta caracterizado predominantemente a la transformación de bienes. Esta industria se encuentra localizada en la ciudad de Chinandega en el Occidente del país.

En 1962, nace industrias GEMINA S.A en un Joint Venture entre General Mills y los accionistas de INA Chinandega a 20 Km del principal puerto de Corinto por donde ingresaban las importaciones al país.

La materia prima utilizada es el trigo que es el elemento principal para la elaboración de harina en la industria importado de Estados Unidos o Europa. Esta constituye un papel importante en la alimentación en nuestro país. El desarrollo de este proceso industrial generara mayor demanda de este producto.

Esta industria satisface tanto el mercado nacional como internacional exportándose actualmente a El Salvador, Honduras y Costa Rica.

Hay dos tipos de trigo que son: North Dark Spring y Soft Red Winter, siendo el primero trigo fuerte y el segundo trigo suave. Su primer molino (actualmente dos) inicia la producción en el mes de Julio de 1963, con las marcas Gold Medal y Bollo fino. Patentizadas respectivamente por la Washburn Crosby y la Sperry Flour Co de Minneapolis, Minesota, Estados Unidos, iniciando el negocio de molienda de trigo para harina panadera en Nicaragua.

Posteriormente aumentó sus líneas de producción con de las siguientes: Harinas para pasta(semolina) del corazón del trigo primavera harina para galletas; suave semi-suave de trigos suaves o blandos y mezclas con semi-duros respectivamente y la mundialmente famosa harina GMKT(Gold Medal Kitchen Tested) para todos los propósitos. Esta ultima es especial para paquetes y reposteria fina en general

Esta empresa esta regida por las normas HACCP que establece los estándares de calidad. En este trabajo se determinarán los intervalos de confianza de las variables humedad, ceniza y proteína que son determinantes en control de calidad y se contrastarán con los intervalos que la empresa ha establecido y determinado con ello si el proceso de producción de la harina se encuentra dentro de los estándares permitidos. Además se establecerán las relaciones funcionales entre las variables para que el Departamento de Control de Calidad pueda realizar análisis retrospectivos y prospectivos en el pronósticos para los futuros periodos aplicando nuestros conocimientos en los problemas que enfrenta la economía del país.

II.- OBJETIVOS

2.1.- OBJETIVO GENERAL.

Analizar las variables: Ceniza, Humedad y Proteína en el proceso de Control de Calidad en la producción de la harina en la Empresa GEMINA S.A. En el período de Agosto a Diciembre 2001.

2.2.- OBJETIVOS ESPECIFICOS

- 1- Describir el proceso de producción de la Harina en la Empresa GEMINA S.A.
- 2- Establecer la medida de asociación y la relación funcional de las variables Ceniza, Humedad y Proteína en el proceso de Control de Calidad de la Empresa GEMINA.
- 3- Determinar los intervalos de confianza de las variables Ceniza, Humedad y proteína.
- 4- Verificar los intervalos de confianza establecidos.

III.- MARCO TEÓRICO

Sobre la calidad del trigo inciden una serie de factores, algunos controlables como el aspecto genético y la fertilidad del suelo, y otro que escapa al control del hombre como es el efecto climático.

Los genetistas deberán continuar esforzándose para mejorar los cultivares tanto en rendimiento como en calidad, el productor prestando atención a la fertilidad del suelo y a la elección del cultivar, y el acopiador realizando un correcto acondicionamiento y clasificación del grano en postcosecha, para lograr satisfacer los requerimientos de los mercados interno y externo de nuestros trigos. Una manera de contribuir a tener trigos de calidad, es a través de la diferenciación y clasificación de la producción triguera.(9)

El proceso de elaboración de la harina involucra una serie de fases que determinan su calidad. Describiremos algunos procedimientos básicos para la elaboración de la misma la extracción de la muestra, el método de análisis de datos los intervalos esperados de las variables a medir. Para ello es importante determinar algunos conceptos y procedimientos que se llevan a cabo en el Proceso de elaboración de la harina. A continuación describiremos brevemente en que consiste y cuales son los elementos básicos a tomar en cuenta en el mismo:

3.1.1.- PROCESO DE ELABORACIÓN DE LA HARINA

Primera Etapa:

El trigo sucio es recibido en las instalaciones de industrias GEMINA vía puerto Corinto este proviene de EE.UU.

Es pesado y descargado en las fosas de descargue pasando seguidamente por la operación de pre-limpiado la cual elimina materias extrañas mayores de una pulgada además de polvo y brosa del mismo cultivo. Esta operación se realiza por una separadora de succión neumática.

Se deposita cada variedad de trigo en silos de almacenamiento separados actualmente contamos con siete silos con capacidad de almacenar 8,800 Tm y la capacidad del sistema de descargue es de 120 Tn/h.

Segunda Etapa:

El trigo es transportado y depositado a los silos de trabajo a través del mismo equipo de descargue, previo a esto el trigo sucio es pesado (Bascula de 400Kg) y pre-limpiado.

Luego es transportado a la maquina de limpieza de trigo. La capacidad del sistema de limpieza es de 6 ton/h.

Pasado a través de la primera maquina de limpieza que es la separadora la cual elimina impurezas de longitud mayor y menor que el trigo, esto lo realiza por mallas o telas metálicas. Es atravesado por una corriente de aire ascendente que elimina las partículas menos densas que el grano como polvo, brosa y granos arrugados. Esta corriente es generada por una turbina de baja presión.

Trasladado por la maquina "Entoleter" la cual elimina el trigo picado, quebrado y polvo, esta es realizada por succión. Pasado a través de una maquina despendedora la cual elimina piedra, vidrios y metales.

Esta maquina aplica el principio de diferencia de densidad para realizar la operación de limpieza. De aquí en adelante el trigo se considera limpio. (1.5-2%) de impureza.

El trigo limpio es mezclado con agua y depositado en los silos de acondicionado contamos con cinco silos con capacidad de 20 Tn cada uno. La cantidad de agua en porcentaje que se le agrega es de (1-5%) de la humedad natural que trae el grano llevando la humedad del grano a un porcentaje de (14-16%) la cual corresponde a la humedad del proceso. El agua se le agrega para ablandar la cáscara del grano y poder extraer el endospermo (harina. El agua que se le agrega es importante por que también determina la humedad de la harina final de empaque.

El trigo limpio es dejado reposar por espacio de 16 horas para que el agua pueda penetrar el grano y posteriormente trasladarlo a la molienda.

Tercera Etapa:

El trigo limpio y acondicionado es transportado a través de un imán para retener partículas presentes en el trigo.

Pasado a través de una maquina ENTOLETER para eliminar polvo y barbilla del trigo.

Pesado para determinar la cantidad de trigo que va a la molienda. La capacidad de molienda es de 5.2 ton/h.

Es transportado a través de un imán para retener partículas ferrosas presentes en el trigo.

Reducido de tamaño (molido). Por medio de 26 pares de rodillos los cuales tienen como objetivo moler el grano y reducir las sémolas del endospermo al tamaño de harina (132 micrones).

El trigo molido es clasificado por maquinas llamadas cernidores las cuales clasifican por granulometría el producto que recibe de la sección de reducción de tamaño. En la sección de clasificado se separa la harina (endospermo) y el salvado (cáscara). Esta sección envía producto al purificado y a la reducción de tamaño nuevamente.

Las sémolas del área de clasificado son transportadas por gravedad a la sección de purificado la cual elimina las partículas de cáscara de trigo que han contaminado las sémolas por medio del principio de diferencia de densidad. Las sémolas purificadas son transportadas a la sección de reducción de tamaño.

Las harinas producidas son recolectadas en dos flujos los cuales son la harina "patente" y harina "clear" estas se diferencian por el color, siendo la harina patente la mas blanca. Las harinas recolectadas pasan a través de imanes para eliminar las partículas ferrosas que pueden estar presentes en la harina.

Son enriquecidas o fortificadas con vitaminas y minerales, también se le agregan aditivos químicos como el bromato de Potasio y blanqueador todos estos son mezclados en maquinas llamadas agitadores.

Son pasadas a través de cernidores de control que eliminan cualquier partícula extraña. Pasan por una bascula para cuantificar la producción de harina. Son pasadas a los entoletter los cuales eliminan cualquier organismo vivo presente en la harina.

Son empacadas y transportadas a la bodega de productos terminados.

3.1.2.- COSTOS DE CALIDAD

Hay que tomar en cuenta los costos de calidad que intervienen en la elaboración de la harina:

Costos preventivos:

En industria GEMINA los problemas de calidad que surgen mas adelante durante la vida del producto se previenen de la siguiente manera: Uno de los problemas es que la harina esta fuera de especificaciones químicas y físicas, una manera para prevenir estos problemas es un Muestreo constante de la harina a la hora del embarque, así como el control riguroso en el molino en cuanto al ajuste de la molienda.

Revisión de los nuevos productos:

Muestra de trigo sucio para análisis físico y químicos; Muestra de trigo sucio almacenado para análisis de investigación con frecuencia diaria; Control de fumigación de los lotes de trigos.

Adiestramiento:

Se realiza seminario cada 6 meses de buenas practicas de mano factura, higiene y seguridad laboral.

Control de procesos:

Chequeo de las maquinas de limpieza para que no sea este enviando a al impureza el trigo cargo de la limpieza, muestreo de la humedad material del trigo, control del factor de humedad.

Obtención y Análisis de los datos de calidad:

Informe diario del laboratorio.

Generación de informe de producción por turno.

Otros, tamizados de control de molienda, chequeo de rendimiento de la molienda, registro por bascula de la cantidad de trigo que se envía a la molienda, chequeo de presencia de bromato de potasio, Muestreo de harina para análisis químicos, Muestreos de trigo para determinación de humedad, de red dog para determinar cenizas.

Costo de Evaluación:

Para determinar la condición del producto desde el punto de vista de la calidad se analiza la materia prima que es el trigo, se le realiza análisis químicos y tipificación del grano.

Inspección y pruebas del material entrante:

Al momento del desembarque del trigo se controla que el pedido este completo y en buen estado.

Inspección y pruebas del producto:

Aquí se da la panificación que consiste en elaborar pan de los diferentes tipos de harina para asegurarse que el producto satisfaga los requisitos de los usuarios.

Conservación de la precisión de los equipos de pruebas:

Registro por bascula de la cantidad de harina producida; ajuste de la reducción de tamaño.

Costos de fallas internas:

Desperdicio:

Cuando el producto sale fuera de especificaciones que no cumple con los análisis físicos químicos y de panificación, la harina no puede salir a la venta y es reprocesada incurriendo en costos de hilo, sacos, mano de obra, energía eléctrica, etc.

Tiempo muerto:

Se da cuando se para la producción ya sea cuando finalizo la molienda, el molino para por atoramiento, el molino para por problemas eléctricos, por problemas en los cernidores, por problemas técnicos, por separaciones.

Costos de fallas externas:

Estos costos se dan pero son aislados.

Devolución de productos o materiales:

Se da cuando el panadero devuelve la harina porque se le mojó o porque no fue el pedido que él hizo por lo cual se reprocesa.

Cargos por garantía:

La única garantía que da GEMINA al consumidor es que la vida útil del producto es de 3 meses. pasado este tiempo la industria no se hace responsable del producto es decir no aceptara reclamos.

3.1.3.- VARIABLES Y ESTÁNDARES DEL PROCESO DE PRODUCCION DE LA HARINA

Las variables que se miden son: humedad, ceniza y proteínas en porcentaje. En que consisten cada una de ellas?

La ceniza: polvo que queda después de una combustión completa, las cenizas de las plantas terrestres contienen potasa y sosa las de las plantas marítimas.

La humedad: cantidad de vapor de agua que contiene la atmósfera e impregna a los cuerpos.

La proteína: compuesto polimétrico de elevado peso moléculas, integrado por una variedad de aminoácidos unidos entre sí por enlaces péptidos.

Cuales son sus estándares?

- 1) Humedad: 13.5% a 14%.
- 2) Cenizas: 0.56% - 0.60%
- 3) Proteínas:
- 4) Climax: 8% - 9%
- 4) Bollo Fino: 9% - 10.9%
- 5) Bollo F. especial: 9% - 9.9%

3.1.4.- PRODUCTOS Y SERVICIOS PRINCIPALES QUE OFRECE GEMINA S.A.

Industrias GEMINA produce principalmente harina de trigo y como producto secundario se obtiene Millrun, lo cual constituye alimento para animales

Dentro de los tipos de harinas que se ofertan los principales son:

- 1) **Harina Gold Medall:** Se caracteriza por poseer proteínas que oscilan entre 12-13% y se usa principalmente en la preparación de pan francés, pan de sal, otros.
- 2) **Harina Bollo Fino:** Posee un contenido proteínico oscila entre un 10-10.9%. es ideal para panadería en general, pan de hamburguesa, hot dog, etc.
- 3) **Harina Bollo Fino Especial:** Su contenido proteínico oscila entre un 9-9.90%. Se usa en la preparación de repostería.
- 4) **Harina Clímax:** Tiene un porcentaje de proteína especial entre 7-8%. Es especial para la preparación de galletas.

Características de los tipos de harina:

- a) Para harinas fuertes: de 12% a 13%.
- b) Para semi suaves: de 10% a 10.9%.
- c) Para suaves: de 8% a 9.9%
- d) Para harina clímax: 7% a 8%.

Materiales Primarios y Auxiliares Secundarios:

La materia prima que se ocupa en esta Industria es el trigo, este es importado de Estados Unidos.

El conjunto de los auxiliares secundarios lo conforman el agua y los ingredientes, estos son:

- 1) **Bromato de potasio:** Este funciona como oxidante, fortaleciendo la estructura del gluten de la harina. A esta se le adicionan 6.44g/quintal.
- 2) **Ácido Ascórbico:** Realiza la misma función que el anterior es un oxidante.
- 3) **Vitaminas:** Las cuales vienen a enriquecer en contenido de la harina. A continuación son:
 - **Tiamina:** Actúa como enzima en las reacciones de oxidación y en el metabolismo de los carbohidratos. Se agregan 6mg por cada 1kg de harina.

- **Riboflavina:** Esta vitamina desempeña un papel importante en el contexto del metabolismo. Se adicionan 4mg por cada 1kg de harina.
 - **Niacina:** El cual es un compuesto de coenzimas de oxidorreducción. A la harina se le adicionan 55mg por cada Kg
 - **Ácido Fólico:** Es una coenzima para transferir carbonos, además interviene en la absorción del hierro. Se agregan 1.5kg por Kg de harina.
- 4) **Hierro :** Se le adiciona 60ppm por 45.35kg de harina.
- 5) **Coroflex (Peroxido de Benzoilo):** Tiene la función de blanqueador. Se agregan 3.7g por 45.35kg de harina.
- 6) **α -Amilasa:** Mejora la actividad enzimática de la harina. Se adicionan 2.29g por quintal.

3.2 TECNICAS ESTADISTICAS PARA EL ANALISIS DE DATOS

Utilizaremos para el análisis de datos:

- Técnicas básicas: Estadísticos descriptivos y métodos gráficos
- Intermedias regresión y correlación lineal para ello abordaremos algunos conceptos básicos.

3.2.1.- EL INTERVALOS DE CONFIANZA.

Un intervalo acotado por dos números se emplea para estimar el valor de un parámetro poblacional. Los valores que limitan este intervalo son estadísticos calculadas a partir de la muestra que están sirviendo como base para la estimación. El nivel de confianza es $1-\alpha$ que la probabilidad de que la muestra por seleccionarse produzca valores limite que se localicen en lados opuestos del parámetro que se estima. Algunas veces el nivel de confianza se llama coeficiente de confianza. El intervalo de confianza se puede expresar como:

$$X \pm Z_{\alpha/2} \cdot Se / \sqrt{n}$$

3.2.2.- ANALISIS DE CORRELACION LINEAL

El objetivo principal del análisis de correlación es medir la intensidad de una relación lineal entre dos variables. Si no existe un cambio definido en los valores de Y conforme aumentan los valores de X, se dice que no hay correlación o que no existe relación entre X e Y. En cambio, si al aumentar X hay una modificación definida de en los valores de Y, entonces existe correlación. En este último caso la correlación es positiva cuando Y tiende a aumentar, y negativa cuando Y decrece. Si tanto los valores de X como los de Y tienden a seguir una dirección recta, existe, una correlación lineal. La precisión del cambio en Y conforme X incrementa su valor, determina la solidez de la correlación lineal.

3.2.3. - ANALISIS DE VARIANZA

El Análisis de Varianza (ANDEVA, o ANOVA, o ANAVA) es uno de los métodos estadísticos más utilizados en las diferentes áreas de la ciencias.

El ANDEVA puede ser definido como una técnica donde la variación total presente en un conjunto de datos es particionada en varias componentes. Asociada con cada una de estas componentes está una fuente de variación específica, tal que en el análisis es posible averiguar la magnitud de la contribución de cada una de esas fuentes para la variación total.

3.2.4.- MODELO LINEAL.

El modelo lineal de la regresión permite abordar el análisis estadístico de que es frecuentemente usado por ejemplo en aquellos trabajos de investigación que están basados en el estudio correlacional de un conjunto de variables, entre las

cuales una de ellas se considera un criterio a predecir (variable dependiente) y las otras predictores de este criterio (variable independiente).

Llamaremos modelo lineal general a la formulación:

$$Y = A + \beta x + u$$

$$A = C$$

Donde Y es un vector de dimensión $n \times 1$ que incluye valores independientes y con la misma varianza σ^2 de una variable respuesta con distribución normal, X es una matriz conocida $n \times p$ de rango $h \leq p$, que contiene los valores de un conjunto de p variable explicativas, x es un vector desconocido de p parámetros, U es un vector $n \times 1$ de variable independientes no observadas $N(0, \sigma)$, H es una matriz de restricciones lineales sobre los p parámetros, con dimensiones, $R \times p$ siendo $r < p$ y rango $(H) = r$ y C es un vector conocido de constantes. El numero de parámetros independientes, es por tanto $K = p - r$. Los dos casos particulares mas importantes de este modelo son:

- El modelo de regresión clásico, donde X incluye valores de variables continuas y su rango es igual a la dimensión del vector β , p , de manera que $X'X$ es no singular. Además, existen como mínimo tantos datos como parámetros, $p \leq n$, y normalmente no existen las restricciones aunque pueden incorporarse fácilmente.
- Los modelos de diseño experimental, donde X incluye valores de variables discretas (en general 0,1 y -1) y su rango es menor que la dimensión del vector de parámetros β siendo por tanto $X'X$ singular.

3.2.5.- ANALISIS DE REGRESIÓN LINEAL MULTIPLE

El análisis de regresión es el conjunto de técnicas que permiten investigar las relaciones entre una variable dependiente, respuesta o endógena (Y) y una variable o mas variables independientes, explicativas o exógena (X_i). En general, una variable Y , depende de muchas otras variables $X_1, X_2, \dots$

X_p , aunque algunas de estas pueden ser no observables o incluso, desconocidas para el investigador. El modelo de regresión pretende medir el efecto de las mas importantes y representa el de las restantes mediante una variable que llamaremos error o perturbación aleatoria.

3.2.6.- DEFINICIÓN DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE

Un modelo de regresión lineal múltiple es de la forma:

$$Y = A + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \epsilon_i \quad \text{donde}$$

A, denota el termino independiente a lo que es lo mismo β_0

$\beta_1, \beta_2, \dots, \beta_K$ son los coeficientes de regresión parcial desconocidos.

ϵ_i Es el termino residual asociado con la i-ésima observación.

3.2.7.- HIPÓTESIS A PRIORI SOBRE LA DISTRIBUCIÓN DE LOS ERRORES

En la aplicación de este modelo supondremos que el termino error verifica las hipótesis siguientes:

- a) Su esperanza es cero.
- b) Su varianza es constante
- c) Las perturbaciones son independientes entre si
- d) Su distribución es Normal
- e) El numero de datos disponibles es como mínimo, $k+1$, esto es $N > (K+1)$
- f) Ninguna de las variables explicativas es una combinación lineal de las demás, es decir las variables X_i son linealmente independientes.
- g) Homocedasticidad: varianzas constantes o igualdad de varianzas
- h) No multicolinealidad: cuando existen altas correlaciones entre las variables independientes, estamos ante el problema de la multicolinealidad.

3.2.8.- LAS HIPÓTESIS RESPECTO A LA PERTURBACIÓN PUEDEN ESCRIBIRSE EN TERMINOS DE LAS VARIABLES RESPUESTAS.

- a) Para cada punto fijado de valores de las X_i la distribución de Y tiene media.

$$E(Y) = A + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- b) La varianza de Y es constante y no depende de los valores de las X_i
Esto es. $\text{Var}(Y) = \sigma^2$

- c) Las variables Y_i son independientes entre si.

- d) Las variables respuesta se distribuyen normalmente.

La estimación de los parámetros la realizaremos a través del METODO DE MINIMOS CUADRADOS. Dicho método consiste en minimizar la suma de cuadrados de los errores.

De acuerdo al criterio de los minimos cuadrados, el valor a minimizar es el producto:

$$\ell' \ell = \ell_1^2 + \ell_2^2 + \dots + \ell_n^2$$

$$\ell' \ell = (Y - Y')' (Y - Y') \text{ y teniendo en cuenta que } Y' = X' \beta$$

$$\text{se tiene } \ell' \ell = (Y - X' \beta)' (Y - X' \beta) = Y' Y + \beta' X' X \beta - 2 \beta' X' Y$$

Para minimizar el error se deriva con respecto al vector del parámetro a estimar

$$\frac{\partial \ell' \ell}{\partial \beta} = 2 X' X \beta - 2 X' Y$$

y se iguala a cero esta derivada:

$$2 X' X \beta - 2 X' Y = 0$$

efectuando transformaciones se obtiene:

$$X' X \beta = X' Y \text{ donde la ecuación resultante es la siguiente:}$$

$\beta = (X'X)^{-1}X'Y$. Que es la expresión matricial de la ecuación normal de regresión. Una vez que ha sido lograda la ecuación de ajuste $Y=X*B$ es necesario pasar a verificar si las variables regresoras en conjunto son capaces de explicar significativamente, el comportamiento de las variables respuestas Y dicha verificación se realiza estableciendo el contraste de que ninguna de las variables explicativas influye en la variable respuesta. Lo que se formula de la siguiente manera.

$$1) H_0 \quad \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 \quad \text{Al menos un } \beta \neq 0$$

El contraste se realiza a través de :

$$F = [VE(x) \quad K] \quad S^2_R$$

De tal forma que si F calculado es mayor que F de la tabla

$$(F_{K, N-K-1, \alpha}) \text{ entonces se rechaza } H_0.$$

3.2.9.- ESPECIFICACION DEL MODELO

El modelo de regresión múltiple que incluye K variables independiente puede expresarse, para un sujeto i , mediante la siguiente ecuación.

$$Y_i = A + \beta_1 X_{i1} + \dots + \beta_J X_{iJ} + \dots + \beta_K X_{iK} + \epsilon_i$$

En donde X_{ij} indica el valor obtenido por el sujeto i en la variable independiente X_j y β_j expresa el coeficiente de regresión parcial, referido a la población asociada a la variable X_j .

La variable ϵ_i presenta el error de predicción para el sujeto i .

La distintas ecuaciones para el conjunto de K sujetos de una población se expresan matricialmente de la siguiente manera:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1J} & \dots & X_{1K} \\ 1 & X_{21} & \dots & X_{2J} & \dots & X_{2K} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & X_{N1} & \dots & X_{NJ} & \dots & X_{NK} \end{pmatrix} * \begin{pmatrix} A \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

$Y_{(N \times 1)}$: Vector de valores de la variable dependiente Y.

$X_{(N \times K)}$: esta matriz contiene N sujetos de la muestra y el conjunto de valores Observados en las variables independientes X_j .

$\beta_{(K \times 1)}$: Vectores que contienen K+1 parámetros a estimar.

$\ell_{(N \times 1)}$: Vector de residuos.

El objetivo del análisis es encontrar el vector de parámetro que sea la mejor estimación del vector desconocido. Para determinar cual es el conjunto de variables independientes que optimizan la predicción de la variable dependiente y eliminar de la ecuación las restantes variables para poder seleccionar el conjunto optimo de variables predictivas el método utilizado es el de inclusión secuencial o STEP-WISE. El ordenador no selecciona las restantes variables dependiente que explican, sino basándose en una mayor reducción de la varianza no explicada por las variables ya incluidas anteriormente en la ecuación.

3.2.10.- PRUEBA “D” DE DURBIN-WATSON

La prueba mas conocida para detectar correlación serial es la desarrollada por los estadísticos Durbin y Watson, mas comúnmente conocida como “d” Durbin-Watson, el cual se define de la siguiente manera.

$$d = \frac{\sum (\ell_t - \ell_{t-1})^2}{\sum \ell_t^2}$$

El cual es simplemente la razón de las sumas de las diferencias al cuadrado de los residuos sucesivos. Una gran ventaja del estadístico (d) consiste en que se basa en los residuos estimados que se calculan automáticamente en el análisis de regresión. Valores grandes del estadístico Durbin-Watson implican auto correlación negativa y valores pequeños indican auto correlación positiva.

PRUEBA D DURBIN-WATSON: REGLA DE DECISION

Hipótesis Nula	Decisión	Si
No existe Autocorrelación positiva	Rechazar	$0 < d < d_L$
No existe auto correlación positiva	No hay decisión	$d_L < d < d_U$
No existe auto correlación Negativa	Rechazar	$4 - d_L < d < 4$
No existe auto correlación negativa	No hay decisión	$4 - d_U < d < 4 - d_L$
No existe auto correlación positiva o negativa	No rechazar	$d_U < d < 4 - d_U$

IV.- MATERIALES Y MÉTODOS

Los datos fueron para el Departamento de Control de Calidad de la Industria GEMINA S.A. que esta bajo la Responsabilidad de la Licenciada Nidia Chui Blandino.

En el proceso productivo la harina se clasifica en tres marcas: Clímax, Bollo Fino y Bollo Fino Especial. Se analizaron los meses de Agosto a Diciembre del 2001 las hojas de registro se muestra en el Anexo 1.

La recolección de los datos se realiza en las 24 horas del día la muestra seleccionada para el control de calidad del proceso de producción se toma bajo un muestreo estratificado y luego se aleatoriza para la toma de la muestra final.

Las variables a estudiar fueron, fecha de análisis, hora de análisis, porcentaje ceniza, porcentaje Humedad y porcentaje de Proteína.

Para el análisis de datos se utilizo el SPSS versión 7.5, para el análisis de datos utilizamos Estadísticos simples (Estadísticos Descriptivo, Métodos gráficos), Estadísticos intermedios y correlación.

V.- RESULTADOS Y DISCUSION.

Los resultados se representarán por las marcas de harina: Clímax, Bollo Fino y Bollo Fino Especial que produce la empresa GEMINA S.A. las variables analizadas fueron: Ceniza, Humedad y Proteína.

Se presentará la distribución de las variables por métodos gráficos y analíticos. La hipótesis que se contrasta es que las variables: Porcentaje de Ceniza, Humedad y Proteínas tienen una distribución normal. Además se presentarán los estadísticos descriptivos de las mismas y la medida de asociación lineal con una primera aproximación de relación funcional.

5.1 DISTRIBUCIÓN DE LAS VARIABLES

A continuación se muestran los gráficos de distribución de las variables por marca:

MARCA CLIMAX:

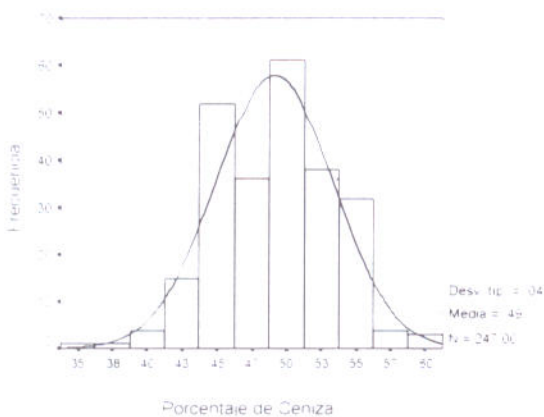


Figura 1. Distribución de la variable Ceniza de la marca Climax en la Empresa GEMINA S.A. de Chinandega Agosto-Diciembre de 2001

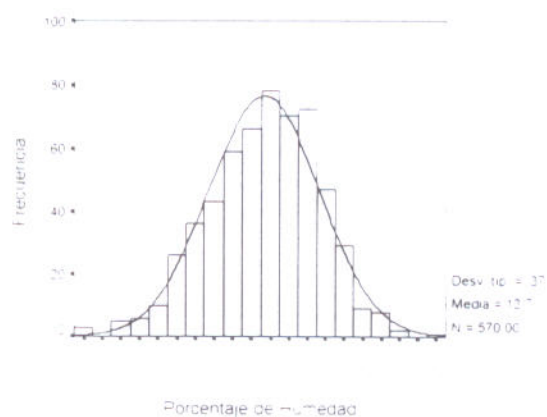


Figura 2. Distribución de la Variable Porcentaje de Humedad de la marca Climax GEMINA S.A. Chinandega Agosto-Dic. del 2001



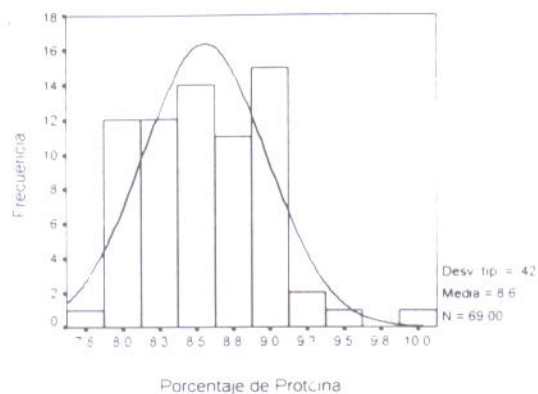


Figura 3. Distribución de la Variable Proteína Marca Climax
en la Empresa GEMINA S.A. Chinandega. Agosto-Dic del 2001.

Las figuras 1-3 nos muestran que la distribución de las variables gráficamente es normal. Realizamos la prueba de Kolmogorov-Smirnov, para comprobar si la hipótesis de Normalidad se cumplía a un nivel de significancia $\alpha=0.05$. No encontramos evidencia suficiente para rechazar la hipótesis de normalidad a ese nivel de significación (Ver Anexo 3, Tabla 1).

MARCA BOLLO FINO:

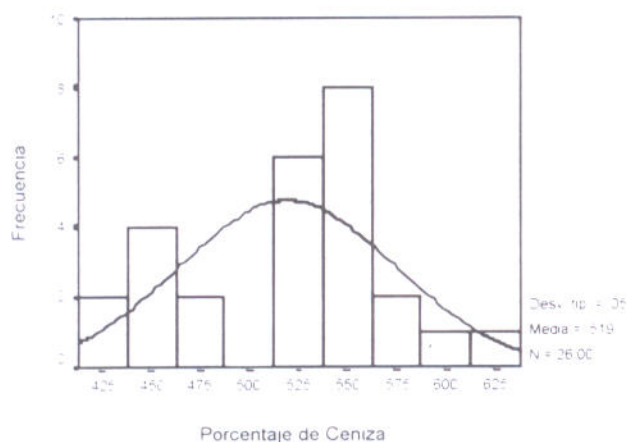


Figura 4. Distribución de la variable Ceniza de la marca Bollo Fino en la
Empresa GEMINA S.A. Chinandega. Agosto-Dic del 2001.

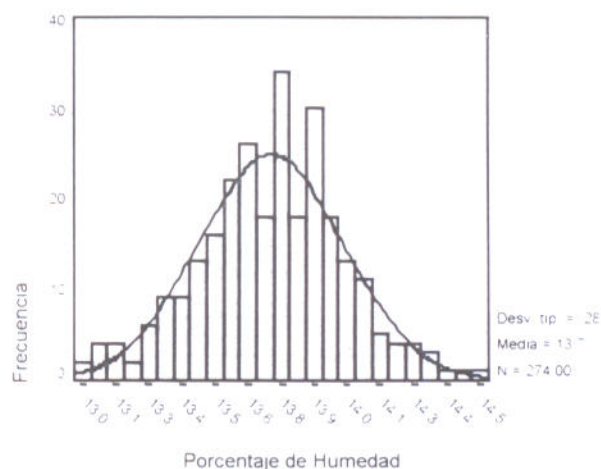


Figura 5. Distribución de la variable Humedad de la marca
Bollo Fino GEMINA S.A. Chinandega. Agosto-Dic del 2001.

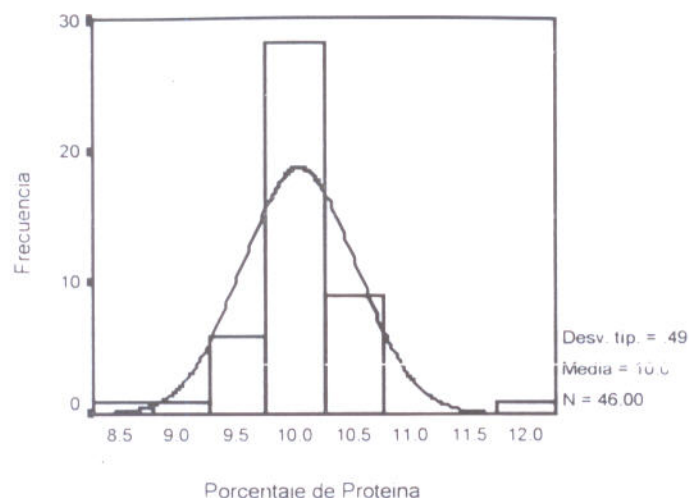


Figura 6 Distribucion de la variable Proteina marca Bollo Fino
Empresa GEMINA S.A. Chinandega. Agosto-Dic del 2001

Las figuras: 4–6 nos muestran que la distribución de las variables gráficamente es normal. Realizamos la prueba de Kolmogorov-Smirnov para comprobar si la hipótesis de normalidad se cumplía a un nivel de significancia $\alpha=0.05$ no encontramos evidencia suficiente para rechazar la hipótesis de normalidad a ese nivel de significación (Ver Anexo 3, Tabla 2).

MARCA BOLLO FINO ESPECIAL:

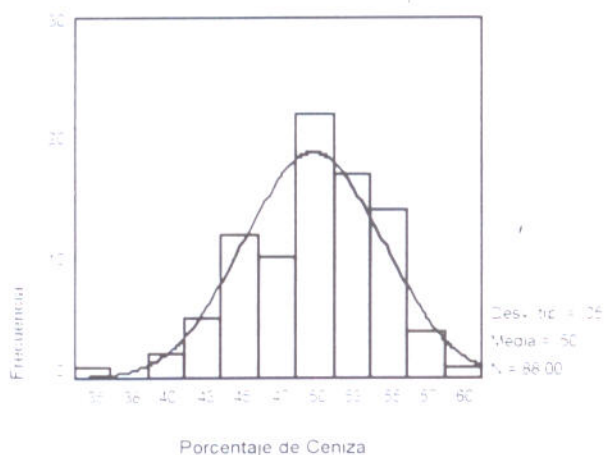


Figura 7 Distribucion de la variable Ceniza marca Bollo Fino Especial
Empresa GEMINA S.A. Chinandega. Agosto-Dic del 2001

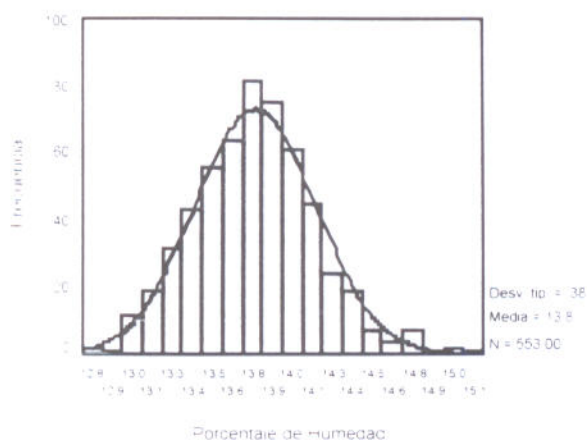


Figura 8 Distribucion de la variable de humedad Bollo Fino Especial
GEMINA S.A. Chinandega. Agosto-Dic del 2001

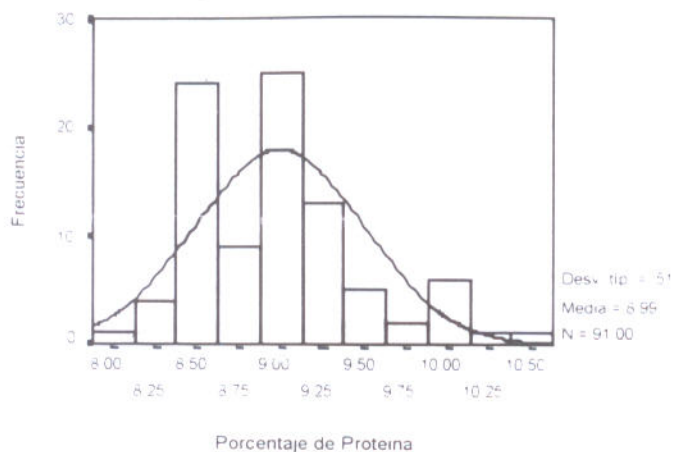


Figura 5: Distribución de la variable Proteína: Bolo Fino Especial
 Empresa: GEMINA S.A. Chinandega. Agosto-Dic del 2001

Las figuras 7-9 nos muestran que la distribución de las variables gráficamente es normal. Realizamos la prueba de Kolmogorov-Smirnov para comprobar si la hipótesis de normalidad se cumplía a un nivel de significancia $\alpha=0.05$. No encontramos evidencia suficiente para rechazar la hipótesis de normalidad a ese nivel de significación (Ver Anexo 3, Tabla 3).

5.2 ESTADÍSTICOS DESCRIPTIVOS

A continuación se presentan los estadísticos descriptivos por variable y por marca.

TIPO DE HARINA: CLIMAX

En la tabla 1 la harina tipo Climax nos muestra las 3 variables: Porcentaje de Ceniza, Humedad y Proteína. Al calcular los intervalos de confianza de la variable Ceniza encontramos que el límite inferior 0.4811 y el límite superior 0.5092. El intervalo establecido por la empresa es de 0.56% a 0.60%, por tanto están por debajo de los límites calculados. Lo que indica es que la variable está bajo control y no supera el intervalo calculado.

La humedad el límite inferior 13.7879% y el superior 14.0242%. El intervalo establecido por la Empresa es de 13% a 14% por tanto están por debajo de los límites calculados. Aunque la Empresa nunca alcanza valores superiores al 14% del Porcentaje de Humedad. Lo que indica es que la variable

esta bajo control y no supera el intervalo calculado. Y la proteína el límite inferior 8.2799 y el superior 8.5884. El intervalo establecido por la Empresa es de 8% a 9% por tanto están por debajo de los límites calculados. Lo que indica que la variable está bajo control y no supera el intervalo calculado.

Se calcula el coeficiente de variación de

$$\text{Ceniza: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.0369}{0.4952} * 100 = 7.45\%$$

$$\text{Humedad: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.3111}{13.9059} * 100 = 2.23\%$$

$$\text{Proteína: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.4055}{8.4341} * 100 = 4.80\%$$

La variable que presenta mayor estabilidad en la media es el Porcentaje de Humedad dado que la media tiene un error de un 2.23% esta es confiable en un 97.77%. Teniendo la mayor variación de la media el Porcentaje de Ceniza siendo confiable en un 92.55%.

Tabla 1. Estadísticos Descriptivos de la Harina Climax Empresa GEMINA S.A Chinandega.
En el periodo Agosto-Diciembre 2001.

			Estadístico	Error tip
Porcentaje de Ceniza	Media		4952	6 852E-03
	Intervalo de confianza para la media al 95%	Limite inferior	4811	
		Limite superior	5092	
	Mediana		4900	
	Varianza		1 362E-03	
	Desv tip		3 690E-02	
	Mínimo		44	
	Máximo		56	
	Rango		12	
Porcentaje de Humedad	Media		13 9059	5 778E-02
	Intervalo de confianza para la media al 95%	Limite inferior	13 7875	
		Limite superior	14 0242	
	Mediana		13 9300	
	Varianza		9 681E-02	
	Desv tip		3111	
	Mínimo		13 33	
	Máximo		14 54	
	Rango		1 21	
Porcentaje de Proteína	Media		8 43417	7 53E-02
	Intervalo de confianza para la media al 95%	Limite inferior	8 27990	
		Limite superior	8 58845	
	Mediana		8 30300	
	Varianza		164	
	Desv tip		40558	
	Mínimo		7 803	
	Máximo		9 604	
	Rango		1 801	

TIPO DE HARINA: BOLLO FINO

En la tabla 2, la harina tipo Bollo Fino nos muestra las 3 variables. Porcentaje de Ceniza, Humedad y proteína. Al calcular los intervalos de confianza de la variable ceniza encontramos que el límite inferior 0.4571 y el límite superior 0.6029. El intervalo establecido por la Empresa es de 0.56%-0.60%, por tanto están por debajo de los límites calculados, lo que indica es que la variable está bajo control y no supera el intervalo calculado.

La Humedad el límite inferior 13.4054 y el límite superior 14.2706 el intervalo establecido por la Empresa es de 13% a 14% por tanto están por debajo de los límites calculados. Lo que indica es que la variable está bajo control y no supera el intervalo calculado.

Y la proteína el límite inferior 9.7221 y el límite superior 10.0458, el intervalo establecido por la Empresa es de 9% a 10.9% por tanto están por debajo de los límites calculados, lo que indica es que la variable está bajo control y no superan los valores en los intervalos calculados.

Se calcula el coeficiente de variación de:

$$\text{Ceniza: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.0587}{0.5300} * 100 = 11.07\%$$

$$\text{Humedad } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.3484}{13.8380} * 100 = 2.51\%$$

$$\text{Proteína } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.1303}{9.884} * 100 = 1.31\%$$

La variable que presenta mayor estabilidad en la media es el Porcentaje de Proteína dado que la media tiene un error de un 1.31% esta es confiable en un 98.69%. Y la que presenta la mayor variación en la media es el Porcentaje de Ceniza siendo confiable solo en 88.93%.

**Tabla 2. Estadísticos Descriptivos de la Harina Bollo Fino Empresa GEMINA S.A
Chinandega. En el periodo Agosto-Diciembre 2001.**

			Estadístico	Error tip
Porcentaje de Ceniza	Media		5300	2.627E-02
	Intervalo de confianza para la media al 95%	Limite inferior	4571	
		Limite superior	6029	
	Mediana		5500	
	Varianza		3.450E-03	
	Desv. tip.		5.874E-02	
	Mínimo		43	
	Máximo		58	
	Rango		15	
Porcentaje de Humedad	Media		13.8380	1558
	Intervalo de confianza para la media al 95%	Limite inferior	13.4054	
		Limite superior	14.2706	
	Mediana		13.9900	
	Varianza		121	
	Desv. tip.		3484	
	Mínimo		13.30	
	Máximo		14.17	
	Rango		87	
Porcentaje de Proteína	Media		9.88400	5.83E-02
	Intervalo de confianza para la media al 95%	Limite inferior	9.72211	
		Limite superior	10.04589	
	Mediana		9.80400	
	Varianza		1.700E-02	
	Desv. tip.		13038	
	Mínimo		9.804	
	Máximo		10.104	
	Rango		300	

TIPO DE HARINA: BOLLO FINO ESPECIAL

En la tabla 3, nos muestra las 3 variables. Porcentaje de ceniza, Humedad y proteína. Al calcular los intervalos de confianza de la variable ceniza, el límite inferior 0.4598 y el límite superior 0.5162. El intervalo establecido por la Empresa es de 0.56% a 0.60%, por tanto es mayor que los límites calculados. Lo que indica que la variable está bajo control y no supera el intervalo calculado.

La humedad el límite inferior 13.7335 y el límite superior 14.1131. El intervalo establecido por la Empresa es de 13% a 14%, por tanto están por debajo de los límites calculados. Lo que indica es que la variable está bajo control y no supera el intervalo.

La proteína el límite inferior 8.4532 y el límite superior 9.0999. El intervalo establecido por la Empresa es de 9% a 9.9% por tanto es mayor que los límites calculados. Lo que indica que la variable está bajo control y no supera el intervalo calculado.

Se calcula el coeficiente de variación del Porcentaje de:

$$\text{Ceniza: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.0508}{0.4880} * 100 = 10.40\%$$

$$\text{Humedad: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.3427}{13.9233} * 100 = 2.46\%$$

$$\text{Proteína: } CV = \frac{S}{\bar{X}} * 100 \text{ entonces } CV = \frac{0.5839}{8.7766} * 100 = 6.65\%$$

La variable que presenta mayor estabilidad en la media es el Porcentaje de Humedad dado que la media tiene un error de un 2.46% esta es confiable en un 97.54%. Teniendo la mayor variación de la media el Porcentaje de Ceniza siendo confiable solo en un 89.6%.

Tabla 3. Estadísticos Descriptivos de la Harina Bollo Fino Especial Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

			Estadístico	Error tip
Porcentaje de Ceniza	Media		4880	1.314E-02
	Intervalo de confianza para la media al 95%	Límite inferior	4598	
		Límite superior	5162	
	Mediana		4900	
	Varianza		2.589E-03	
	Desv. tip		5.088E-02	
	Mínimo		40	
	Máximo		55	
	Rango		15	
Porcentaje de Humedad	Media		13.9233	8.849E-02
	Intervalo de confianza para la media al 95%	Límite inferior	13.7335	
		Límite superior	14.1131	
			13.9220	
	Varianza		117	
	Desv. tip		3427	
	Mínimo		13.38	
	Máximo		14.49	
	Rango		1.11	
Porcentaje de Proteína	Media		8.77660	15077
	Intervalo de confianza para la media al 95%	Límite inferior	8.45324	
		Límite superior	9.09996	
	Mediana		8.60300	
	Varianza		341	
	Desv. tip		58391	
	Mínimo		8.303	
	Máximo		10.504	
	Rango		2.201	

5.3.- ANÁLISIS DE REGRESIÓN Y CORRELACION

5.3.1.- CORRELACIONES SIMPLES

Tabla 4. Análisis de correlación de la Harina Climax Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

		Porcentaje de Ceniza	Porcentaje de Humedad	Porcentaje de Proteína
Correlación de Pearson	Porcentaje de Ceniza	1.000	- 161*	578**
	Porcentaje de Humedad	- 161*	1.000	- 556**
	Porcentaje de Proteína	578**	- 556**	1.000
Sig (bilateral)	Porcentaje de Ceniza		012	001
	Porcentaje de Humedad	012		000
	Porcentaje de Proteína	001	000	

* La correlación es significativa al nivel 0.05 (bilateral)

** La correlación es significativa al nivel 0.01 (bilateral)

Tabla 5. Análisis de correlación de la Harina Bollo Fino. Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

		Porcentaje de Ceniza	Porcentaje de Humedad	Porcentaje de Proteína
Correlación de Pearson	Porcentaje de Ceniza	1.000	- 198	116
	Porcentaje de Humedad	- 198	1.000	- 405
	Porcentaje de Proteína	116	- 405	1.000
Sig (bilateral)	Porcentaje de Ceniza		376	827
	Porcentaje de Humedad	376		135
	Porcentaje de Proteína	827	135	

Tabla 6. Análisis de correlación de la Harina Bollo Fino Especial. Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

		Porcentaje de Ceniza	Porcentaje de Humedad	Porcentaje de Proteína
Correlación de Pearson	Porcentaje de Ceniza	1.000	-.293*	.177
	Porcentaje de Humedad	-.293*	1.000	-.323*
	Porcentaje de Proteína	.177	-.323*	1.000
Sig (bilateral)	Porcentaje de Ceniza		.010	.496
	Porcentaje de Humedad	.010		.030
	Porcentaje de Proteína	.496	.030	

* La correlación es significativa al nivel 0.05 (bilateral)

En las tablas. 4.5 y 6 los presentan las correlaciones simples de los tipos de harina Climax, Bollo Fino y Bollo Fino Especial con las tres variables en estudio. La tendencia en los tres tipos de harina es que la Humedad presenta una correlación negativa con el porcentaje de Ceniza y Proteína.

Las hipótesis a contrastar son:

H_0 : No hay relación entre Ceniza y Humedad

H_1 : Hay relación entre Ceniza y Humedad

A un nivel de significancia $\alpha = 0.05$, fijado para los análisis estadísticos encontramos evidencia suficiente para rechazar la hipótesis nula para las marca Climax y Bollo Fino Especial (Tabla 4 y 6). Sin embargo el tipo Bollo Fino no se rechaza la hipótesis nula a ese nivel de significación (Tabla 5). Por tanto estadísticamente para estas variables la correlación es no significativa.

Hipótesis a contrastar:

H_0 : No hay relación entre el Porcentaje de Ceniza y Proteína

H_1 : Hay relación entre Porcentaje de Ceniza y Proteína

Al nivel de significancia α fijado, encontramos evidencia suficiente para rechazar la hipótesis nula para las marca Climax (Tabla 4). Para el tipo Bollo Fino y Bollo

Fino Especial (Tabla 4 y 6) no se rechaza la hipótesis nula a este nivel de significación (Tabla 5 y 6).

Las hipótesis a contrastar son:

H_0 : No hay relación entre Proteína y Humedad

H_1 : Hay relación entre Proteína y Humedad

Al nivel de significancia α descrito anteriormente, no encontramos evidencia suficiente para rechazar la hipótesis nula para las marca Clímax y esta correlación es altamente significativa (Tabla 4). Para el tipo Bollo Fino Especial se rechaza la hipótesis nula (Tabla 6) esta correlación es significativa. Sin embargo el tipo Bollo Fino no se rechaza la hipótesis nula a ese nivel de significación (Tabla 5). Por tanto estadísticamente para estas variables la correlación es no significativa.

5.3.2.- ANALISIS DE REGRESION MÚLTIPLE

El análisis de regresión múltiple se realizó por tipo de harina. Las variables fueron:

Y: Porcentaje de Proteína

X₁: Porcentaje de Ceniza

X₂: Porcentaje de Humedad

Los análisis de varianza se presentarán por tipo de Harina.

TIPO CLIMAX:

Bajo la hipótesis:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Encontramos evidencia suficiente a un nivel de significancia $\alpha = 0.05$, para rechazar la hipótesis nula. Por lo tanto los parámetros del modelo de regresión son significativos (Tabla 8 y 9)

El modelo de regresión estimada es:

$$Y_i = 14.297 + 4.179 X_1 - 0.570 X_2 + \varepsilon_i$$

Que nos indica que el porcentaje de Proteína incrementara 4.17 si el Porcentaje de Ceniza incrementa en una unidad y se mantiene fija el Porcentaje de Humedad además el porcentaje de Proteína disminuirá 0.57 cuando la humedad incremente en una unidad manteniendo fija el porcentaje de Ceniza. Por lo tanto hay una relación directa entre el Porcentaje de Proteína y Ceniza e inversamente proporcional al porcentaje de Humedad.

Al validar el modelo encontramos que no podemos rechazar la hipótesis de autocorrelación en los errores ni la de homogeneidad en la varianza. El modelo cumple los supuestos del Análisis de Varianza por lo que no se invalidan ninguna de las pruebas del mismo. (Tabla 7)

Tabla 7. Resumen del modelo de la Harina Climax Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Modelo	R	R cuadrado	R cuadrado corregida	Error tip de la estimación	Durbin-Watson
1	677 ^a	458	416	30991	2.201

^a Variables predictoras (Constante). Porcentaje de Humedad. Porcentaje de Ceniza

^b Variable dependiente. Porcentaje de Proteína

Tabla 8. Análisis de varianza del modelo de regresión de la Harina Climax Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	2.109	2	1.054	10.977	000 ^a
	Residual	2.497	26	9.605E-02		
	Total	4.606	28			

^a Variables predictoras (Constante). Porcentaje de Humedad. Porcentaje de Ceniza

^b Variable dependiente. Porcentaje de Proteína

Tabla 9. Coeficientes del modelo de regresión de la Harina Climax Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Variables del modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%		Estadísticos de colinealidad	
	B	Error tip	Beta			Limite inferior	Limite superior	Tolerancia	FIV
(Constante)	14.297	3.219		4.441	.000	7.680	20.915		
Porcentaje de Ceniza	4.179	1.705	.380	2.451	.021	.674	7.685	.866	1.154
Porcentaje de Humedad	-.570	.202	-.438	-2.821	.009	-.986	-.155	.866	1.154

^a Variable dependiente. Porcentaje de Proteína

TIPO BOLLO FINO:

Bajo la hipótesis:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

No encontramos evidencia suficiente para rechazar la hipótesis nula por lo tanto el modelo es no significativo.

$$Y_i = 10.198 - 1.946X_1 + 0.0518X_2 + \varepsilon_i$$

La hipótesis de independencia de los errores no se cumple igual que la de homogeneidad en la varianza.

Tabla 10. Resumen del modelo de la Harina Bollo Fino. Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Modelo	R	R cuadrado	R cuadrado corregido	Error típ. de la estimación	Durbin-Watson
1	.892 ^a	.796	.592	8.3282E-02	1.646

a. Variables predictoras (Constante), Porcentaje de Humedad, Porcentaje de Ceniza

b. Variable dependiente: Porcentaje de Proteína

Tabla 11. Análisis de varianza del modelo de regresión de la Harina Bollo Fino Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	5.413E-02	2	2.706E-02	3.902	.204 ^a
	Residual	1.387E-02	2	6.936E-03		
	Total	6.800E-02	4			

a. Variables predictoras (Constante), Porcentaje de Humedad, Porcentaje de Ceniza

b. Variable dependiente: Porcentaje de Proteína

Tabla 12. Coeficientes del modelo de regresión de la Harina Climax Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Variables en el modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%		Estadísticos de colinealidad	
	B	Error tip	Beta			Límite inferior	Límite superior	Tolerancia	FIV
(Constante)	10.198	1.711		5.962	.027	2.838	17.558		
Porcentaje de Ceniza	-.1946	.709	-.876	-2.743	.111	-4.998	1.107	.999	1.001
Porcentaje de Humedad	5.182E-02	.120	.138	.433	.707	-.463	.566	.999	1.001

a Variable dependiente: Porcentaje de Proteína

TIPO BOLLO FINO ESPECIAL:

Bajo la hipótesis:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

No encontramos evidencia suficiente para rechazar la hipótesis nula por lo tanto el modelo es no significativo.

$$Y_i = 16.303 - 0.507X_1 - 0.523X_2 + \varepsilon_i$$

La hipótesis de independencia de los errores no se cumple igual que la de homogeneidad en la varianza

Tabla 13. Resumen del modelo de la Harina Bollo Fino Especial. Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Modelo	R	R cuadrado	R cuadrado corregido	Error tip. de la estimación	Durbin-Watson
1	.283 ^a	.080	-.073	60494	1.688

a Variables predictoras: (Constante), Porcentaje de Humedad, Porcentaje de Ceniza

b Variable dependiente: Porcentaje de Proteína

Tabla 14. Análisis de varianza del modelo de regresión de la Harina Bollo Fino Especial Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	382	2	191	522	606 ^a
	Residual	4 391	12	366		
	Total	4 773	14			

a. Variables predictoras: (Constante), Porcentaje de Humedad, Porcentaje de Ceniza

b. Variable dependiente: Porcentaje de Proteína

Tabla 15. Coeficientes del modelo de regresión de la Harina Bollo Fino Especial Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%		Estadísticos de colinealidad	
	B	Error tip.	Beta			Límite inferior	Límite superior	Tolerancia	FIV
(Constante)	16.303	9.426		1.730	.109	-4.235	36.840		
Porcentaje de Ceniza	-.507	3.945	-.044	-.129	.900	-9.104	8.089	.649	1.542
Porcentaje de Humedad	-.523	.586	-.307	-.893	.390	-1.799	.753	.649	1.542

a. Variable dependiente: Porcentaje de Proteína

En los análisis de varianza de los modelos de regresión múltiple para los diferentes tipos de harina no se puede establecer la tendencia en la relación de las variables. Es conveniente seguir realizando estudios de regresión múltiples para otros periodos para valorar la tendencia y determinar si es necesario incluir otras variables exógenas dentro del modelo.

Es claro que el Porcentaje de Humedad es clave para mantener a la harina fuera de focos de contaminación de bacterias u hongos que hacen que esta se deteriore con la elevación de la temperatura.

En los Análisis de la Tabla 13, el modelo es no significativo. Por tanto no podemos establecer un modelo de regresión lineal múltiple.

VI.- CONCLUSIONES

- ♦ Al calcular los intervalos de confianza de las variables: Porcentaje de Humedad, Ceniza y Proteína por Tipo de Harina producida: Clímax, Bollo Fino y Bollo Fino Especial con los datos que registra el departamento de Control de Calidad de la Empresa GEMINA S.A., no encontramos evidencia suficiente para decir que estos no cumplen con los establecidos por la empresa. Y al evaluar la distribución de las mismas encontramos que se distribuyen normal supuesto básico para la construcción de los intervalos de confianza. Por tanto estas variables están bajo control.
- ♦ Al calcular el coeficiente de variación por variables y por tipo de harina encontramos que en el Tipo Clímax y Bollo Fino Especial la variable Porcentaje de Humedad es la que presenta menor coeficiente de variación por lo tanto la media de estas variables es más confiable hasta en un 97.77%. La variable que mayor variación tiene para los tres tipos de Harina es el Porcentaje de Ceniza y esta llega a ser confiable hasta en un 92.55%. En el único caso que el Porcentaje de Proteína presenta mayor estabilidad en la media es en el caso del Tipo Bollo Fino que es confiable en 98.69% en el Tipo Clímax es confiable en 95.2% y Bollo Fino Especial en 93.35%.
- ♦ Al realizar el análisis de correlación simple entre las variables: Porcentaje de Humedad, Ceniza y Proteína se pudo observar que las correlaciones simples de los tipos de harina Clímax, Bollo Fino y Bollo Fino Especial existe una tendencia de correlación negativa con el porcentaje de Ceniza y Proteína. La relación entre Ceniza y Humedad, a un nivel de significancia $\alpha = 0.05$, fijado para los análisis estadísticos encontramos evidencia suficiente para decir que estas variables tienen una correlación significativa. Para las marca Clímax encontramos que el Porcentaje de Ceniza y Proteína tienen una

correlación significativa. Sin embargo para los tipos Bollo Fino y Bollo Fino Especial estas tienen una correlación no significativa. La asociación entre las variables Porcentaje de Humedad y Proteína en el tipo de Harina Climax esta correlación es altamente significativa. Para el tipo Bollo Fino Especial esta correlación es significativa. Sin embargo el tipo Bollo Fino la correlación es no significativa.

- ♦ Al realizar el análisis de regresión lineal múltiple donde la variable endógena fue Porcentaje de Proteína y las exógenas fueron: Porcentaje de Humedad y Ceniza solo en el tipo de harina Climax se encontró que el modelo era significativo y al validarlo cumplía con los supuestos básicos. En los otros tipos de Harina el modelo fue no significativo. No hay una tendencia clara en la evaluación de los signos del modelo estimado.

VII.- RECOMENDACIONES

- ♦ Sería conveniente realizar estos estudios en las dos épocas del año para validar los intervalos de confianza establecidos por la Empresa GEMINA S.A.
- ♦ Es recomendable seguir estudios en el área de la modelización para que la empresa cuente con modelos de pronósticos en las dos épocas del año.
- ♦ Es importante potencializar la anuencia de los directivos de la Empresa GEMINA S.A. en el departamento de Matemática y Estadística y con ello poner en práctica los conocimientos adquiridos en la carrera de Estadística.

VIII.- BIBLIOGRAFÍA

- 1) Berascout Grijalva Jorge Efrain. Importancia de la agroindustria en Centro-América
- 2) Ishikawa Dr. Kaoru. Guía de Control de Calidad.
- 3) Anuarios del Banco Central de Nicaragua.
- 4) D'León Farley Proyecto para la implementación de producción más limpia en Industria GEMINA S.A.
- 5) Envista a Lic. Chui Blandino Nidia. Tecnóloga de Alimentos en la Empresa GEMINA S.A.
- 6) Johson Robert Estadística Elemental. S.A. de C.V. 1988.
- 7) Riba M.D-J.M Doménech Métodos Estadísticos Modelo Lineal de Regresión. Barcelona 1985.
- 8) Peña Sánchez de Rivera Daniel. Estadística Modelos y Métodos Fundamentos S.A. Madrid 1986 a 1998
- 9) [http://www.aaprotrigo.org/quintocongreso/CONDICIONES AMBIENTALES Y GENETICAS CAL. TRIGO- 2001 \(1\). htm](http://www.aaprotrigo.org/quintocongreso/CONDICIONES AMBIENTALES Y GENETICAS CAL. TRIGO- 2001 (1). htm)
- 10) http://rafaela.inta.gov.ar/publicaciones/trigo2001/misc94_4.htm

IX.- ANEXOS

FECHA:

MOLINO:

ANALISIS DE HARINAS DE PRODUCCION

14% HUM. 14% HUM.

MARCA	HORA	% HUMEDAD	% CENIZA	% PROTEINA	ppm BROMATO	ppm A. ASC.	ppm HIERRO	FALLING
					VISUAL	VISUAL	VISUAL	NUMBER, seg
	07:00 a.m.							
	08:00 a.m.							
	09:00 a.m.							
	10:00 a.m.							
	11:00 a.m.							
	12:00 p.m.							
	01:00 p.m.							
	02:00 p.m.							
	03:00 p.m.							
	04:00 p.m.							
	05:00 p.m.							
	06:00 p.m.							

HUMEDADES DE TRIGOS

HORA	1 BK	1BK	TRIGO	RED DOG	RED DOG	CARGA	% GLUTEN	
	ENTERO	MOLINO	SECO	%HUMEDAD	% CENIZA			
07:00 a.m.								
08:00 a.m.								
09:00 a.m.								
10:00 a.m.								
11:00 a.m.								
12:00 p.m.								
01:00 p.m.								
02:00 p.m.								
03:00 p.m.								
04:00 p.m.								
05:00 p.m.								
06:00 p.m.								

PROMEDIOS

HUMEDAD:

CENIZA:

PROTEINA

FN:

GLUTEN:

OBSERVACIONES:

MOLINO:

ANALISIS DE HARINAS DE PRODUCCION

14% HUM. 14% HUM.

MARCA	HORA	% HUMEDAD	% CENIZA	% PROTEINA	ppm BROMATO	ppm A. ASC.	ppm HIERRO	FALLING
					VISUAL	VISUAL	VISUAL	NUMBER, seg.
	07:00 p.m.							
	08:00 p.m.							
	09:00 p.m.							
	10:00 p.m.							
	11:00 p.m.							
	12:00 a.m.							
	01:00 a.m.							
	02:00 a.m.							
	03:00 a.m.							
	04:00 a.m.							
	05:00 a.m.							
	06:00 a.m.							

HUMEDADES DE TRIGOS

HORA	1 BK ENTERO	1 BK MOLIDO	TRIGO SECO	RED DOG %HUMEDAD	RED DOG % CENIZA	CARGA	% GLUTEN
07:00 p.m.							
08:00 p.m.							
09:00 p.m.							
10:00 p.m.							
11:00 p.m.							
12:00 a.m.							
01:00 a.m.							
02:00 a.m.							
03:00 a.m.							
04:00 a.m.							
05:00 a.m.							
06:00 a.m.							

PROMEDIOS

HUMEDAD:

CENIZA:

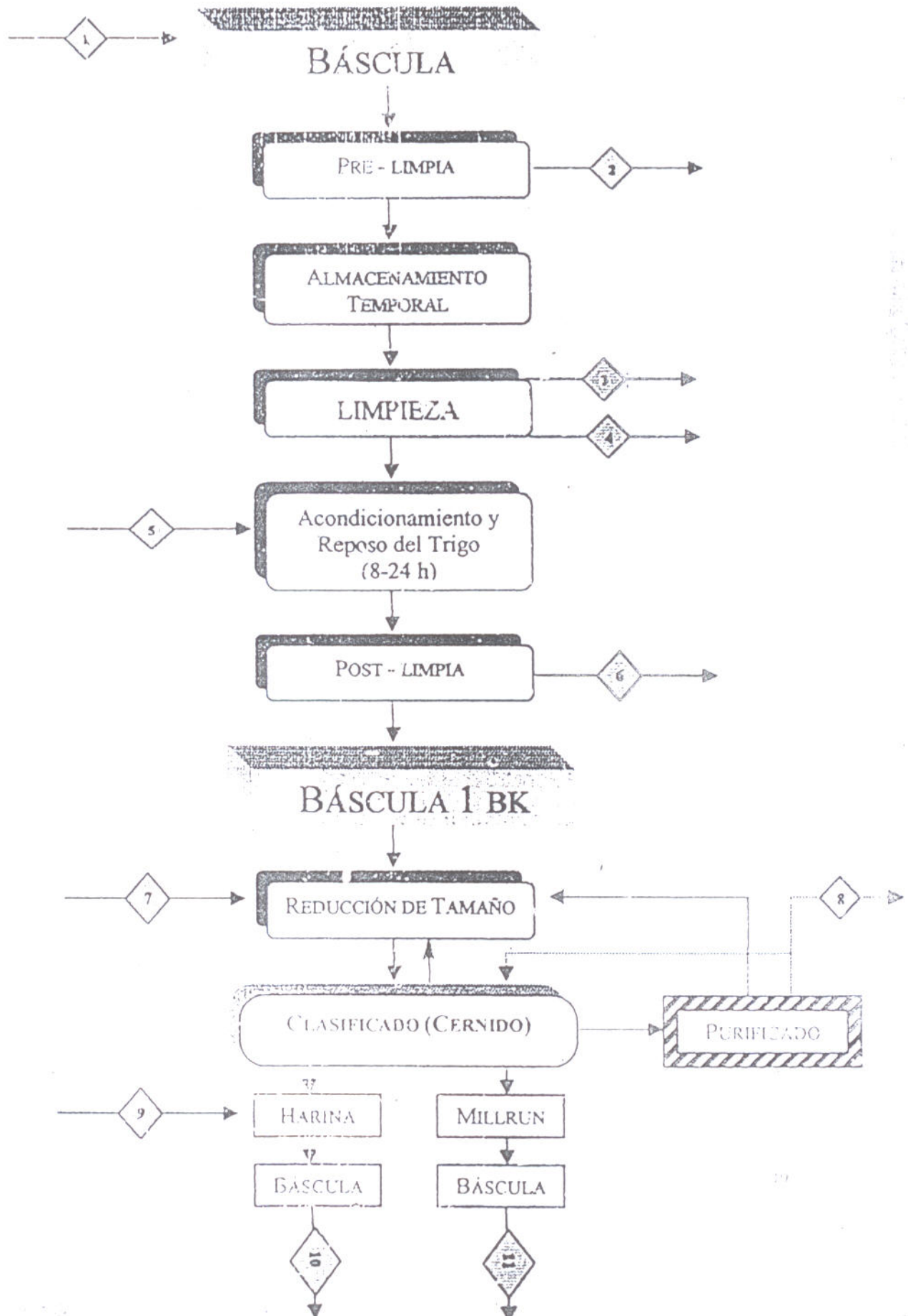
PROTEINA.

FN:

GLUTEN:

OBSERVACIONES:

Diagrama de Bloque



ANEXO 3:

PRUEBAS DE NORMALIDAD

Tabla 1. Prueba de Kolmogorov-Smirnov para la Harina Climax Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

		Porcentaje de Ceniza	Porcentaje de Humedad	Porcentaje de Proteína
N		247	570	69
Parámetros normales ^{a,b}	Media	4911	13 7102	8 55836
	Desviación típica	4 252E-02	3720	42050
Diferencias más extremas	Absoluta	076	039	105
	Positiva	076	024	105
	Negativa	- 063	- 039	- 082
Z de Kolmogorov-Smirnov		1 194	938	869
Sig. asintót. (bilateral)		115	342	436

a. La distribución de contraste es la Normal

b. Se han calculado a partir de los datos.

Tabla 2. Prueba de Kolmogorov-Smirnov para la Harina Bollo Fino Empresa GEMINA S.A Chinandega. En el periodo Agosto-Diciembre 2001.

		Porcentaje de Ceniza	Porcentaje de Humedad	Porcentaje de Proteína
N		26	274	46
Parámetros normales ^{a,b}	Media	5192	13 7114	10 02789
	Desviación típica	5 433E-02	2750	49095
Diferencias más extremas	Absoluta	198	045	150
	Positiva	125	034	144
	Negativa	- 198	- 045	- 150
Z de Kolmogorov-Smirnov		1 009	739	1 019
Sig. asintót. (bilateral)		260	646	250

a. La distribución de contraste es la Normal

b. Se han calculado a partir de los datos

Tabla 3. Prueba de Kolmogorov-Smirnov para la harina Bollo Fino Especial. Empresa GEMINA S.A. Chinandega. En el período Agosto-Diciembre 2001

		Porcentaje de Ceniza	Porcentaje de Humedad	Porcentaje de Proteína
N		88	553	91
Parámetros normales ^{a,b}	Media	4993	13 7674	8 98707
	Desviación típica	4 675E-02	3778	50543
Diferencias más extremas	Absoluta	090	032	147
	Positiva	067	032	147
	Negativa	- 090	- 015	- 081
Z de Kolmogorov-Smirnov		842	764	1 403
Sig. asintót. (bilateral)		478	604	051

^a La distribución de contraste es la Normal

^b Se han calculado a partir de los datos

