

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**INSTITUTO DE CIENCIAS MATEMÁTICAS**

**PROYECTO DE GRADUACIÓN**

**PREVIO A LA OBTENCIÓN DEL TÍTULO DE:**

**“MAGÍSTER EN GESTIÓN DE LA PRODUCTIVIDAD  
Y LA CALIDAD”**

**TEMA**

**“ANÁLISIS DE REDUCCIÓN DE PÉRDIDA EN LA FABRICACIÓN DE CUBIERTAS  
ONDULADAS”**

**AUTORES**

**JOSE DAVID ZAMBRANO MUÑOZ**

**KARLA CARMEN MIRANDA RAMOS**

**Guayaquil- Ecuador**

**AÑO**

**2012**

# **DEDICATORIA**

A nuestros padres, hermanas y hermanos

## **AGRADECIMIENTO**

A nuestros padres y hermanos que con su amor y paciencia han sido fuente de inspiración para alcanzar nuestros sueños.

Además queremos agradecer a todas y cada una de las personas que nos dedicaron su tiempo y aportaron sus ideas para el desarrollo de nuestro trabajo.

## **DECLARACIÓN EXPRESA**

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Graduación, nos corresponde exclusivamente; y el Patrimonio Intelectual del mismo, corresponde exclusivamente al **ICM (Instituto de Ciencias Matemáticas)** de la Escuela Superior Politécnica del Litoral.

---

JOSE DAVID ZAMBRANO MUÑOZ

---

KARLA CARMEN MIRANDA RAMOS

# TRIBUNAL DE GRADUACIÓN

---

M.Sc. JOHN RAMÍREZ FIGUEROA  
DIRECTOR DEL PROYECTO

---

PhD. FRANCISCO VERA ALCÍVAR  
PRESIDENTE DEL TRIBUNAL

---

MPC. MIRIAM RAMOS BARBERÁN  
VOCAL DEL TRIBUNAL

# **AUTORES DEL PROYECTO**

---

JOSE DAVID ZAMBRANO MUÑOZ

---

KARLA CARMEN MIRANDA RAMOS

# CONTENIDO

	Pág.
<b>LISTA DE ABREVIATURAS .....</b>	<b>xiii</b>
<b>INTRODUCCIÓN .....</b>	<b>xiv</b>
<b>PLANTEAMIENTO DEL PROBLEMA.....</b>	<b>xv</b>
<b>JUSTIFICACIÓN.....</b>	<b>xvi</b>
<b>OBJETIVO GENERAL.....</b>	<b>xvii</b>
<b>OBJETIVOS ESPECÍFICOS .....</b>	<b>xvii</b>
<b>CAPÍTULO I .....</b>	<b>1</b>
<b>1. MARCO DE REFERENCIA .....</b>	<b>1</b>
1.1 DESCRIPCIÓN DEL PROCESO .....	1
1.2 CONTROL ESTADÍSTICO DE PROCESO.....	4
1.3 FUNDAMENTOS DE ESTADÍSTICA.....	5
1.3.1 DEFINICIÓN.....	5
1.3.2. TIPOS DE ESTADÍSTICA .....	5
1.3.2.1. ESTADÍSTICA DESCRIPTIVA .....	5
1.3.2.2. ESTADÍSTICA INFERENCIAL .....	6
1.3.3. RECOPIACIÓN DE DATOS.....	6
1.3.3.1. POBLACIONES Y MUESTRAS .....	6
1.3.3.2. TIPOS DE MUESTRAS .....	7
1.3.3.3. TIPOS DE DATOS .....	9
1.3.3.4 FUENTES DE DATOS .....	9
1.3.4. PRESENTACIÓN DE LOS DATOS .....	9
1.3.4.1 TABLAS DE FRECUENCIA .....	9
1.3.4.2. DIAGRAMA DE SECTORES .....	10
1.3.4.3. HISTOGRAMA .....	10

1.3.4.4.	DIAGRAMA DE CAJAS.....	11
1.3.4.5.	GRÁFICA DE BARRAS .....	12
1.3.4.6.	DIAGRAMA DE DISPERSIÓN .....	13
1.3.4.7.	SIMETRÍA DE DATOS.....	13
<b>1.3.5.</b>	<b>MEDIDAS DE TENDENCIA CENTRAL .....</b>	<b>14</b>
1.3.5.1.	MEDIA .....	15
1.3.5.2.	MEDIANA .....	15
1.3.5.3.	MODA .....	15
<b>1.3.6.</b>	<b>MEDIDAS DE VARIABILIDAD O DISPERSIÓN.....</b>	<b>16</b>
1.3.6.1	RANGO .....	16
1.3.6.2	VARIANZA .....	16
1.3.6.3.	DESVIACIÓN ESTÁNDAR.....	16
<b>1.3.7.</b>	<b>NIVEL DE SIGNIFICANCIA.....</b>	<b>17</b>
<b>1.3.8.</b>	<b>TIPOS DE ERRORES .....</b>	<b>17</b>
<b>1.3.9.</b>	<b>ANÁLISIS PARAMÉTRICO.....</b>	<b>18</b>
<b>1.3.10.</b>	<b>TIPOS DE ANOVA.....</b>	<b>19</b>
1.3.10.1	ANÁLISIS SIMPLE DE VARIANZA .....	19
1.3.10.2.	ANÁLISIS DOBLE DE VARIANZA .....	22
<b>1.3.11.</b>	<b>DISTRIBUCION "F" FISHER.....</b>	<b>23</b>
<b>1.3.12.</b>	<b>R CONSOLE (SOFTWARE) .....</b>	<b>25</b>
<b>CAPÍTULO II.....</b>		<b>26</b>
<b>2. ANÁLISIS DE LA INFORMACIÓN .....</b>		<b>26</b>
<b>2.1.</b>	<b>DISEÑO EXPERIMENTAL.....</b>	<b>33</b>
<b>2.2.</b>	<b>DESARROLLO .....</b>	<b>35</b>
2.2.1.	MODELO 1: ANÁLISIS DE INFLUENCIA POR VARIABLE TURNO. ....	35
2.2.2.	MODELO 2: ANALISIS DE INFLUENCIA POR VARIABLE OPERARIO .....	41
2.2.3	MODELO 3: ANÁLISIS DE VARIABLES SIN INTERACCIÓN.....	47
2.2.4	MODELO 4: ANÁLISIS DE VARIABLES CON INTERACCIÓN.....	50



<b>CAPÍTULO III .....</b>	<b>54</b>
<b>3. PROPUESTA DE DISEÑO DE MEJORA .....</b>	<b>54</b>
3.1 PLANTEAMIENTO DE TURNOS DE TRABAJO .....	55
3.2 PROGRAMACIÓN SEMANAL DE TURNOS DE TRABAJO .....	55
<b>CONCLUSIONES .....</b>	<b>59</b>
<b>RECOMENDACIONES .....</b>	<b>60</b>
<b>BIBLIOGRAFÍA .....</b>	<b>61</b>

# ÍNDICE DE TABLAS

	Pág.
<i>Tabla I-1</i> Diagrama de flujo Paneladora 4 Bradbury Guayaquil _____	3
<i>Tabla I-2</i> Tipos de muestreos _____	8
<i>Tabla I-3</i> Análisis simple de la varianza _____	21
<i>Tabla II-1</i> Datos Paneladora#4 _____	27
<i>Tabla II-2</i> Datos atípicos Paneladora# 4 _____	30
<i>Tabla II-3</i> Datos depurados Paneladora# 4 _____	30
<i>Tabla II-4</i> Histograma: datos depurados de pérdidas Paneladora# 4 _____	31
<i>Tabla II-5</i> Análisis de varianza de un factor (pérdidas –turno) _____	35
<i>Tabla II-6</i> Análisis de varianza de un factor (pérdidas –turno) _____	36
<i>Tabla II-7</i> Análisis de medias entre turnos (pérdidas Paneladora #4) _____	37
<i>Tabla II-8</i> Tabla de medias por turno (pérdidas Paneladora 4) _____	38
<i>Tabla II-9</i> Comparación entre <b>RSS0</b> y <b>RSS1</b> (pérdidas Paneladora 4) _____	39
<i>Tabla II-10</i> Análisis de varianza de un solo factor (pérdidas-operario) _____	41
<i>Tabla II-11</i> Análisis de varianza de un solo factor (pérdidas-operario) _____	42
<i>Tabla II-12</i> Análisis de medias entre operarios (pérdidas Paneladora 4) _____	43
<i>Tabla II-13</i> Medias de pérdidas por operarios (pérdidas Paneladora 4) _____	44
<i>Tabla II-14</i> Comparación entre RSS0 y RSS2 (pérdidas Paneladora 4) _____	45
<i>Tabla II-15</i> Análisis sin interacción entre variables turno, pérdida y operario _____	48
<i>Tabla II-16</i> Comparación entre modelo aditivo y modelo con interacción (pérdidas Paneladora 4) _____	50
<i>Tabla II-17</i> Análisis de la interacción entre variables turno, pérdidas y operario _____	51
<i>Tabla II-18</i> Porcentaje de pérdida para cada operario por turno (pérdidas Paneladora 4) _____	52
<i>Tabla III-1</i> Pérdidas promedio de operadores entre turnos (Paneladora 4) _____	54
<i>Tabla III-2</i> Cuadro de operadores (Paneladora 4) _____	55
<i>Tabla III-3</i> Propuesta semana 1 (Paneladora 4) _____	56
<i>Tabla III-4</i> Propuesta semana 2 (Paneladora 4) _____	56

	<b>Pág.</b>
<i>Tabla III-5 Propuesta semana 3 (Paneladora 4)</i> _____	57
<i>Tabla III-6 Propuesta semana 4 (Paneladora 4)</i> _____	57
<i>Tabla III-7 Promedio de pérdida mensual estimado</i> _____	58

# ÍNDICE DE GRÁFICOS

	<b>Pág.</b>
<i>Gráfico I.1 Diagrama de recorrido Paneladora 4 Bradbury Guayaquil</i> _____	2
<i>Gráfico I.2 Diagrama de sectores</i> _____	10
<i>Gráfico I.3 Histograma</i> _____	11
<i>Gráfico I.4 Diagrama de cajas</i> _____	12
<i>Gráfico I.5 Gráfica de barras</i> _____	12
<i>Gráfico I.6 Diagrama de dispersión</i> _____	13
<i>Gráfico I.7 Medidas representativas de un conjunto de datos estadísticos</i> _____	14
<i>Gráfico II.1 Histograma de pérdidas Paneladora#4</i> _____	28
<i>Gráfico II.2 Diagrama de cajas de pérdidas Paneladora#4</i> _____	29
<i>Gráfico II.3 Diagrama de cajas: datos depurados de pérdidas Paneladora# 4</i> _____	32
<i>Gráfico II.4 Diagrama de cajas: comparativo de pérdidas Paneladora# 4</i> _____	33
<i>Gráfico II.5 Diagrama de cajas: medias comparativas de turnos Paneladora 4</i> _____	40
<i>Gráfico II.6 Diagrama de cajas: medias comparativas de operarios (Paneladora 4)</i> ____	47
<i>Gráfico II.7 Pronóstico sin Interacción entre turnos y operarios (Paneladora 4)</i> _____	49
<i>Gráfico II.8 Pronósticos con interacción entre turnos y operarios (Paneladora 4)</i> _____	52

## LISTA DE ABREVIATURAS

<b>INEN.</b>	Instituto Ecuatoriano de Normalización
<b>P-VALUE.</b>	Estadístico valor p
<b>ANOVA.</b>	Análisis de Varianza
<b>MS.</b>	Mean Square
<b>OUTLIERS.</b>	Valores atípicos
<b>SDC.</b>	Suma de diferencias cuadráticas
<b>DMC.</b>	Diferencias cuadráticas medias
<b>F.</b>	Fisher
<b>PSE.</b>	Problem Solving Environments
<b>STD. ERROR.</b>	Error estándar
<b>T- VALUE.</b>	Valor T
<b>RSE.</b>	Residual standard error
<b>RSS.</b>	Residual standard square
<b>SQ.</b>	Squared

## INTRODUCCIÓN

Las organizaciones del siglo XXI se encuentran inmersas en una incansable búsqueda por mejorar sus productos, sus servicios, a través de la administración de sus métodos aplicados en las áreas específicas que se desarrollan los negocios. Sin embargo, el tema de calidad y reducción de desperdicios es un tema importante que abarca toda la organización que se compromete a mejorar, involucrando a todos los elementos que forman parte de ella.

También es necesario establecer claramente los estándares de calidad, y así poder cubrir todos los aspectos relacionados al sistema de calidad.

Para dar efecto a la implantación de mejoras, es necesario que todos tengan los conocimientos requeridos para conocer las exigencias del proceso.

Este proyecto se basó en el mejoramiento continuo por medio del análisis de reducción de pérdidas en la fabricación, observando el comportamiento de las variables que intervienen.

Se demostrará una mejora continua del proceso, partiendo del análisis estadístico de datos de producción recopilados en 36 meses.

La idea de este trabajo de investigación es el mejoramiento en el proceso de fabricación de cubiertas. El requisito más importante para lograr el éxito consiste en obtener un alto grado de compromiso adquirido por la alta dirección, así como también del resto del personal que labora en la empresa.

## **PLANTEAMIENTO DEL PROBLEMA**

El proceso de producción de cubiertas galvanizadas, tenía establecido un límite de hasta 2% de desperdicios conformados por cubiertas galvanizadas fuera de las especificaciones establecidas en la Norma INEN 2221:2008 [9], por presencia de manchas y material considerado como chatarra; sin embargo, en un período de tres años el proceso ha presentado fluctuaciones importantes en su porcentaje de pérdidas que en su mayoría no cumple con los límites establecidos, lo que genera una reducción en el rendimiento del proceso.

Por tal razón se propone hacer un estudio del efecto de la calidad de materiales, empleo del procedimiento, habilidades de los operarios y la incidencia de los horarios de producción, sobre el índice de desperdicio.

Cada vez resulta más importante el tema de la mejora continua y eficiencia en los resultados, para asegurar que los procesos estén diseñados para cumplir con los requisitos y así lograr ser más competitivos en el mercado en cual se participa.

Por medio de este proyecto, se pretende presentar una propuesta de solución para este tipo de problemas, aplicando como herramientas el análisis estadístico descriptivo del proceso, y el mejoramiento continuo de los procesos.

A través de las siguientes interrogantes podemos guiar nuestra investigación:

¿Se conocen históricamente problemas directos que influyen en la generación de las pérdidas?

¿Qué tan variable es el proceso productivo y operativo en la fabricación de cubiertas onduladas?

¿Cuáles son las variables que intervienen dentro del proceso?

¿Se conoce el grado de relación de dependencia del índice de desperdicio con el resto de factores del proceso?

## JUSTIFICACIÓN

Debido a que la gestión de calidad es un medio para brindar confianza en el producto o servicio adquirido, actualmente en todo el mundo existe una creciente orientación hacia la globalización que conlleva al mejoramiento continuo de los procesos.

El Ecuador por cumplimiento de la ley del Sistema Ecuatoriano de la Calidad según el Art. 3 declara como política de estado “*La demostración y la promoción de la calidad, en los ámbitos público y privado, como un factor fundamental y prioritario de la productividad, competitividad y del desarrollo nacional*”, los organismos tales como el Ministerio de Productividad e Industria, Instituto de Normalización Ecuatoriana y el Organismo de Acreditación Ecuatoriano entre otros, colaboran con el propósito de proteger al consumidor final mediante el control y verificación de los sistemas de calidad para asegurar que los productos que adquiridos o fabricados cumplen con los requisitos especificados en una Normativa o Reglamento, es por eso que la empresa en estudio considera orientarse hacia la eficiencia y productividad del proceso mediante la optimización del uso de sus recursos y el incremento de la satisfacción de sus clientes.

Luego de analizar a la empresa objeto de este proyecto, está claro que existe un alto grado de interés por parte de la Dirección de la Empresa en temas de optimización de procesos mediante la reducción de pérdidas, por lo cual se ha decidido que en ella se desarrollará este proyecto.

La metodología aplicada se basará en probar mediante el análisis estadístico descriptivo, el comportamiento de los procesos y los factores que intervienen, así como determinar cuáles son los principales problemas que inciden en la pérdida y la evaluación para el planteamiento de mejoras.



## **OBJETIVO GENERAL**

Reducir pérdidas en la fabricación de cubiertas onduladas metálicas, determinando las variables que inciden directamente; además plantear una mejora que nos ayude a incrementar la eficiencia y productividad para alcanzar el nivel óptimo del proceso.

## **OBJETIVOS ESPECÍFICOS**

Disminuir el porcentaje de desperdicios en el proceso de producción de cubiertas onduladas galvanizadas en un período de un año.

### **ESTRATEGIAS:**

- Efectuar el análisis de causa de desperdicio en el proceso de fabricación de cubiertas onduladas metálicas.
- Determinar los principales problemas que indiquen en el proceso de transformación de la materia prima.
- Probar mediante un análisis estadístico descriptivo el comportamiento del proceso y los factores que intervienen.
- Analizar, evaluar y realizar el planteamiento de mejoras de optimización del proceso mediante la disminución de las pérdidas en la fabricación.

## CAPÍTULO I

### 1. MARCO DE REFERENCIA

#### 1.1 DESCRIPCIÓN DEL PROCESO

La planta de producción está conformada por varias líneas conformadoras en frío de paneles metálicos, entre las que se encuentra la línea paneladora # 4 dedicada a la fabricación de cubiertas metálicas onduladas.

Para la consecución o elaboración del producto final es necesaria la intervención de personas calificadas y de experiencia, quienes mediante métodos y procedimientos establecidos cumplen con los requerimientos de calidad establecidos por el cliente.

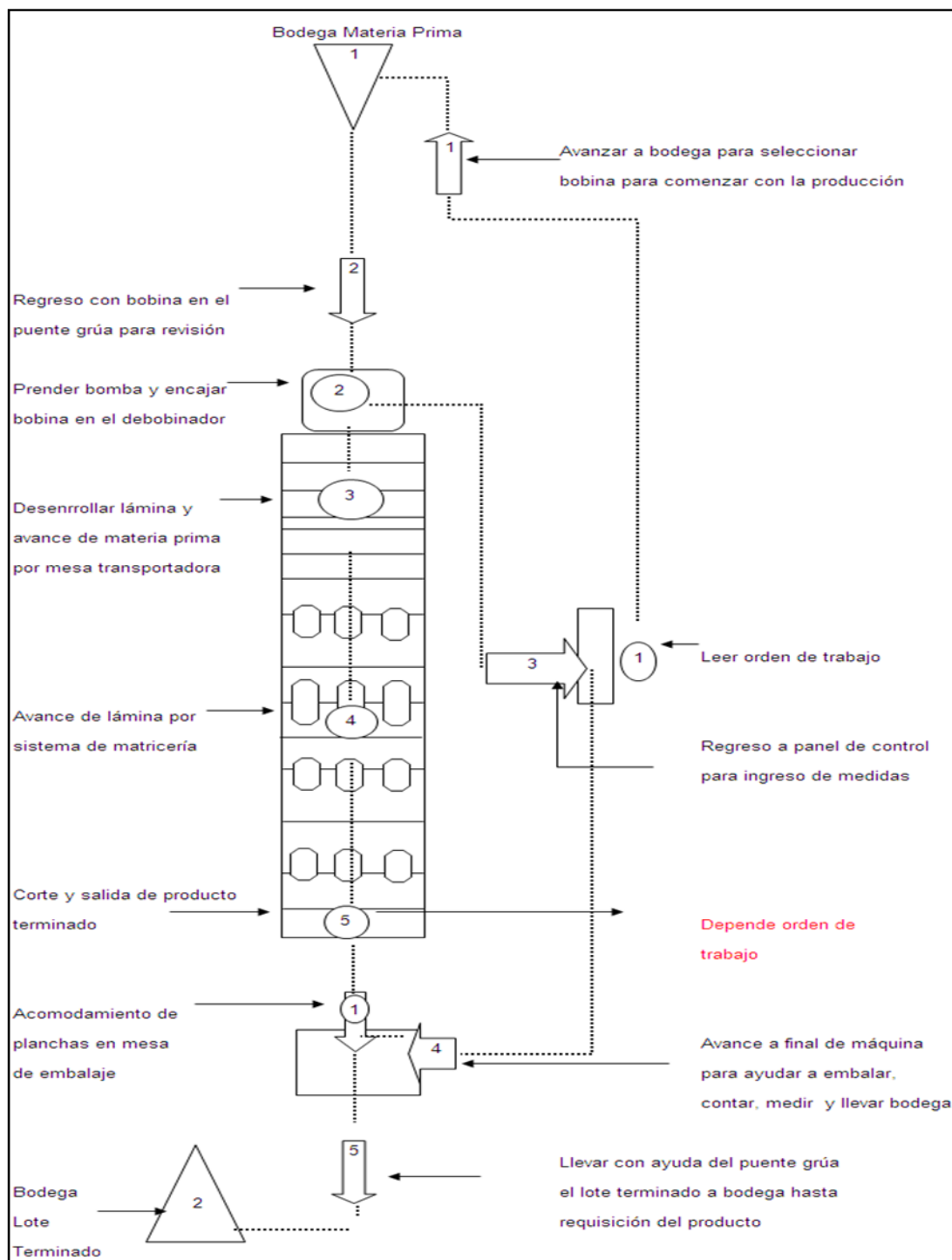
Paneladora 4 Bradbury Guayaquil, está conformada por los siguientes dispositivos.

**Debobinador:** es una máquina que requiere como insumo bobinas metálicas, cuya función es desenrollar láminas, mediante la rotación del brazo mecánico donde se coloca la bobina.

**Roll Former:** es una máquina a la cual ingresa la lámina metálica para ser conformada en frío por medio de matricerías, que las conforman rodillos ubicados a ciertas distancias de manera que se estructure la lámina en un panel metálico.

**Cizalla:** es una máquina en la cual ingresa la lámina metálica, cuya función es realizar un corte transversal a una cierta longitud, medida automáticamente por un sensor, el cual es programado por el operador.

**Apilador:** esta sección está conformada por mesas metálicas móviles donde el operador y ayudante realizan el acomodamiento del panel conformado hasta ser transportados a su ubicación final.



*Gráfico I.1 Diagrama de recorrido Paneladora 4 Bradbury Guayaquil  
Elaborado por: Los autores*

DIAGRAMA DE FLUJO DE PROCESO							
Objeto del Diagrama:		Paneladora Bradbury Guayaquil			Diagrama :		1
El diagrama empieza en :					Diagrama de metodos:		
Fecha:		22-Mar-12			Elaborado por:		
Producto:		Novazinc			Nombre OP:		J. Triviño
Orden de Trabajo:		28300 und; long. 3600			Nombre AY:		Ayovi
Dist. M	Tiempo sg.	Simbolo	Descripción Proceso	Dist. M	Tiempo sg.	Simbolo	Descripción del Proceso
	145,2sg	1	Leer orden de trabajo antes de empezar la producción	-			
	137,4sg	1	Caminar a bodega por materia prima para inicio de rproceso	-	3,0sg	4	Tiempo espera para comienzo de trabajo
	12,0sg	2	Localización visual de la bobina	-	3,0sg	8	Avance a panel de control del operador
	123,0sg	1	Seleccionar bobina para inicio de proceso	-	28,2sg	10	Codificar medidas y especificaciones para arranque de proceso
	124,2sg	1	Inspección de bobina que cumpla con los parametros necesarios para inicio de trabajo	-	5,4sg	11	Encender paneladora para avance de laminas en proceso
	79,8sg	1	Espera por puente grua	-	4,2sg	9	Avance de lamina por mesa transportadors
	35,4sg	2	Ir donde esta ubicado el puente grua	-	10,2sg	2	Avance de plancha por el sistema de matriceria
	72,6sg	3	Regresar con puente grua para transportar bobina	-	1600,0sg	12	Corte y salida de producto terminado
	245,4sg	4	Trasportar bobina al area de trabajo para inicioo de proceso	-	3,6sg	13	Medición de producto terminado
	190,2sg	3	Descargar bobina al area de trabajo para prepararla antes del montaje al rebobinador	-	13,2sg	14	Conteo de número de planchas
	120,6sg	4	Montar bobina al puente grua	-	125,4sg	15	Embalaje de monto de planchas
	122,4sg	5	Tranportar bobina al rebobinador	-	-	5	Espera puente grua ocupado
	3,6sg	6	Caminar hacia bomba para encender el rebobinador	-	6,6sg	10	Caminar donde esta ubicado el puente grua
	2,4sg	5	Encender bomba para abrir mordazas	-	34,2sg	11	Regresar con el puente grua donde la bobina
	12,6sg	6	Encajar bobina a las mordazas del rebobinador	-	71,4sg	16	Preparar producto terminado para montaje puente grua
	34,2sg	7	Retiro de puente grua	-	84,0sg	17	Embarque de producto terminado puente grua
	79,2sg	8	Desenrollar lamina de la bobina poner mesa trasportadora para iniciar proceso hasta salida de mesa trasportadora	-	79,8sg	12	Camino con el lote a bodega
	66,6sg	2	Retraso para corregir falla	-	120,0sg	18	Desembarco de producto terminado a bodega
	6,0sg	9	Acomodar lamina	-	Indeterminado	1	Almaoerar hasta que haya requisición
	10,8sg	3	Tiempo de para para inicio de proceso	-			
	4,2sg	7	Avance a panel de control del operador	-			

*Tabla I-1 Diagrama de flujo Paneladora 4 Bradbury Guayaquil  
Elaborado por: Los autores*

### **Descripción de la Materia prima.**

La principal materia prima utilizada para la fabricación de las cubiertas onduladas son las bobinas de acero, las cuales tienen varios espesores que van desde 0.20 mm hasta 0.60 mm. El peso de las bobinas es de 6 toneladas en promedio.

### **Galvanizado.**

El galvanizado es una práctica común en todo el mundo en lo referente a la industria de la construcción, utilizado comúnmente para proteger las piezas de acero que van a ser expuestas a condiciones ambientales adversas por un largo tiempo.

Se denomina galvanizado por inmersión en caliente al proceso mediante el cual se obtiene recubrimientos sobre aceros u otros materiales féreos por inmersión en un baño de zinc fundido.

Los sistemas que se utilizan para evitar la corrosión del hierro y el acero son esenciales para la utilización económica de estos metales como materiales de construcción. La prescripción en el proyecto de un buen sistema de protección supone una economía considerable, ya que se elaboran gastos de conservación y evitan las interrupciones del servicio, además de aumentar la vida útil del equipo, complemento a instalación.

## **1.2 CONTROL ESTADÍSTICO DE PROCESO**

El control de los procesos es una actividad de vital importancia para cualquier organización, ya que le permite visualizar su posición respecto a la planificación inicial de sus actividades y en función de estas tomar las decisiones pertinentes a cada caso.

Con base en esta necesidad se han diseñado una gran cantidad de herramientas que permiten el control y análisis de los procesos, con el fin único de optimizarlos y asegurar la satisfacción del cliente.

Las técnicas de control de procesos se pueden clasificar en dos grupos, las herramientas clásicas y las nuevas herramientas. El primer grupo está compuesto por Diagramas de Flujo de Proceso, Diagrama Causa - Efecto, Diagrama de Pareto, Diagrama de Correlación o Dispersión, Histogramas, Cuestionario Crítico y Hoja de Seguimiento.

Las nuevas herramientas son aquellas diseñadas para evaluar situaciones que las herramientas clásicas no pueden detectar. [1]

## **1.3 FUNDAMENTOS DE ESTADÍSTICA**

### **1.3.1 DEFINICIÓN**

La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones.

Cuando coloquialmente se habla de estadística, se suele pensar en una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia de nuestro entorno, ya que hoy día es casi imposible que cualquier medio de difusión, periódico, radio, televisión, etc., no nos aborde diariamente con cualquier tipo de información estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, etc.

Sólo cuando nos adentramos en un mundo más específico como es el campo de la investigación de las Ciencias Sociales: Medicina, Biología, Psicología, etc., empezamos a percibir que la Estadística no sólo es algo más, sino que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordadas desde la perspectiva de las leyes deterministas. Podríamos, desde un punto de vista más amplio, definir la estadística como la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre.[2]

### **1.3.2. TIPOS DE ESTADÍSTICA**

Por lo general el estudio de la estadística se divide en dos categorías:

Estadística descriptiva y Estadística Inferencial.

#### **1.3.2.1. ESTADÍSTICA DESCRIPTIVA**

Se dedica a los métodos de recolección, descripción, visualización y resumen de datos originados a partir de los fenómenos en estudio. Los datos pueden ser resumidos numéricamente o gráficamente. Ejemplos básicos de parámetros estadísticos son: la media y la desviación estándar. Algunos ejemplos gráficos son: histograma, pirámide poblacional, clústeres, etc.[2]

### **1.3.2.2. ESTADÍSTICA INFERENCIAL**

Se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Se usa para modelar patrones en los datos y extraer inferencias acerca de la población bajo estudio.

Estas inferencias pueden tomar la forma de respuestas a preguntas si/no (prueba de hipótesis), estimaciones de características numéricas (estimación), pronósticos de futuras observaciones, descripciones de asociación (correlación) o modelamiento de relaciones entre variables (análisis de regresión). Otras técnicas de modelamiento incluyen anova, series de tiempo y minería de datos.[2]

### **1.3.3. RECOPIACIÓN DE DATOS**

#### **1.3.3.1. POBLACIONES Y MUESTRAS**

En los estudios estadísticos es crucial identificar la población (el grupo, sujetos u objetos sometidos a estudio).

Una **población** es el conjunto completo de individuos o elementos de interés.

Un **censo** es un intento de medir todos los elementos de una población de interés.

Una **muestra** es un subconjunto seleccionado de la población.

En la mayor parte de las investigaciones estadísticas, los censos son bastante costosos y difíciles, o incluso imposibles. Bajo las anteriores condiciones es necesario seleccionar una muestra que es un subconjunto de la población elegida para el análisis.[2]

### **1.3.3.2. TIPOS DE MUESTRAS**

Existen dos tipos de muestras: aleatorias y no aleatorias. La **muestra aleatoria** simple resulta cuando se seleccionan  $n$  elementos de una población, de manera que todas las combinaciones posibles de  $n$  elementos de la población tienen igual probabilidad de ser elegidas. Si algunos de los elementos de la población tienen mayor posibilidad de ser elegidos que otros, esto constituye una **muestra no aleatoria**.

Si es posible obtener una muestra aleatoria simple es mejor, ya que por lo general éste es el mejor método. Es fácil realizarlo, es fácil entenderlo y no se puede aducir sesgo en la muestra una vez que se hizo la selección.

Sin embargo, no siempre es posible realizar un muestreo aleatorio simple. Para que este método se pueda usar, debe disponerse de una lista de la población.

Una **muestra sistemática** es aquella para la que se selecciona cada  $n$ -ésimo elemento de la población.

En un **muestreo aleatorio** estratificado la población se divide en estratos o subgrupos significativos y se obtiene una muestra aleatoria de cada uno.

El **muestreo por Conglomerados** implica la selección aleatoria de grupos de conglomerados de elementos como parte de la muestra.[2]



<b>MÉTODOS</b>		<b>PROCEDIMIENTOS</b>
<b>Aleatorios</b>		
<b>Simple</b>		Los elementos se eligen al azar
<b>Sistemático</b>		Se elige cada n-ésimo elemento de una secuencia de la población
<b>Estratificado</b>		La población se separa en subgrupos significativos antes del muestreo
<b>Por conglomerados</b>		Se elige grupos o conglomerados de elementos de la población para la muestra
<b>No aleatorios</b>		
<b>Subjetivo</b>		Se usa el juicio del investigador para decidir qué elementos de la población serán incluidos en la muestra
<b>Por conveniencia</b>		Se eligen los elementos más convenientes para la muestra
<b>Por cuota</b>		Se les asigna una cuota a aquellos que seleccionan elementos de la muestra

*Tabla I-2 Tipos de muestreos  
Fuente: Métodos Estadísticos Aplicados por Pedro Sánchez*

### **1.3.3.3. TIPOS DE DATOS**

Existen dos tipos básicos de variables: 1) cualitativas y 2) cuantitativas. Cuando la característica que se estudia es de naturaleza no numérica, recibe el nombre de **variable cualitativa o atributo**. Algunos ejemplos de variable cualitativa son el género, la filiación religiosa, tipo de automóvil que se posee. Cuando los datos son de naturaleza cualitativa importa la cualidad o proporción que caen dentro de cada categoría.

Cuando la variable que se estudia aparece de forma numérica, la variable se denomina variable cuantitativa. Ejemplos de variables cuantitativas son el saldo de su cuenta de cheques, las edades de los presidentes de la compañía, el tiempo vida de la batería de un automóvil.

Las variables cuantitativas pueden ser discretas o continuas. Las **variables discretas** adoptan solo ciertos valores y existen vacíos entre ellos. Ejemplos de variables discretas son el número de camas de una casa, el número de estudiantes en cada sección del curso.

Las observaciones de una **variable continua** toman cualquier valor dentro de un intervalo específico. Ejemplos de una variable continua son la presión del aire en una llanta y el peso de un cargamento de tomates. Por lo general, las variables continuas son el resultado de las mediciones.[2]

### **1.3.3.4 FUENTES DE DATOS**

Los datos necesarios para elaborar un análisis estadístico o bien se encuentran disponibles o deben recogerse. Los datos que se encuentran disponibles se denominan **datos secundarios** y los datos que deben recogerse se llaman **datos primarios**.[2]

## **1.3.4. PRESENTACIÓN DE LOS DATOS**

### **1.3.4.1 TABLAS DE FRECUENCIA**

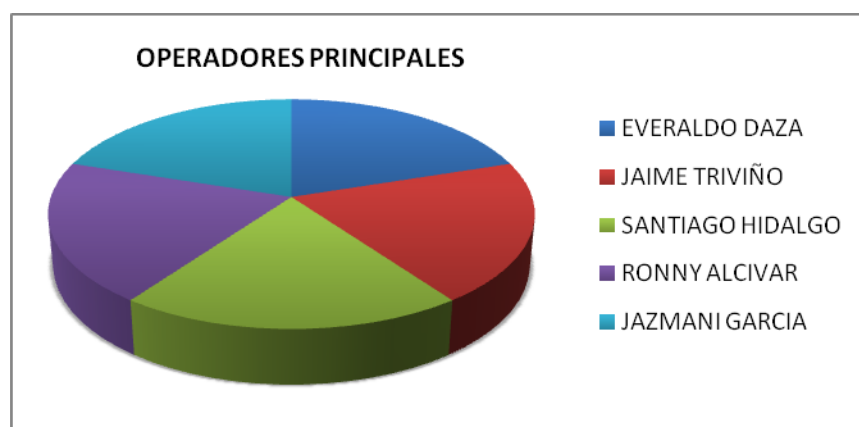
Es una tabla resumen en la que se disponen los datos divididos en grupos ordenados numéricamente, denominados clases o categorías. El número de datos u observaciones que pertenecen a determinada clase se llama **frecuencia de clase**, el punto medio de cada clase o

categoría se llama **marca de clase** y la longitud de una clase se conoce como **intervalo de clase**. La **frecuencia absoluta** es el número o cantidad de observaciones iguales o semejantes que se encuentran comprendidas dentro de un determinado intervalo de clase.

La **frecuencia relativa** es el cociente que resulta de dividir la frecuencia absoluta de una clase para la suma total de frecuencias de todas las clases de una tabla de frecuencias. [1]

### 1.3.4.2. DIAGRAMA DE SECTORES

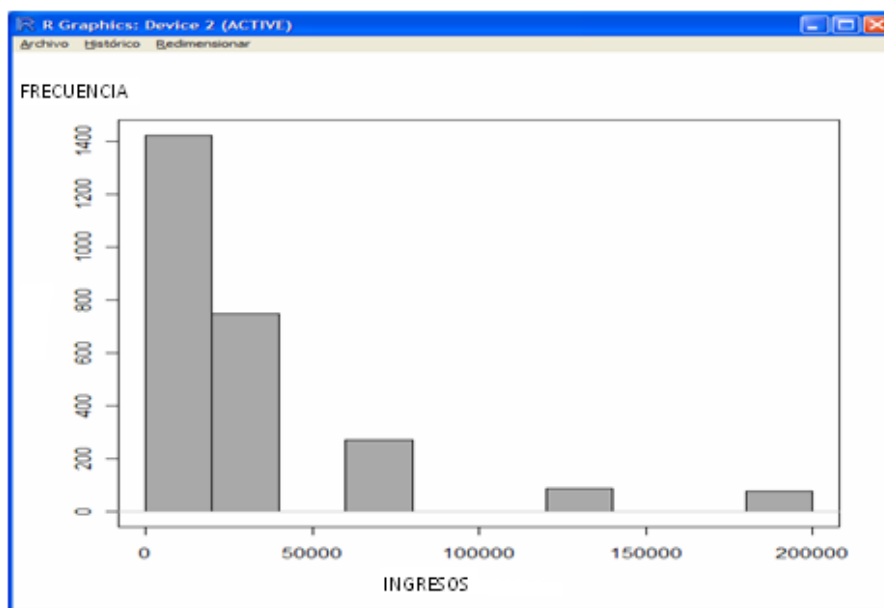
Esta representación gráfica consiste en dividir un círculo en tantos sectores circulares como modalidades presente el carácter cualitativo, asignando un ángulo central a cada sector circular proporcional, consiguiendo de esta manera un sector con área proporcional.[1]



*Gráfico 1.2 Diagrama de sectores  
Elaborado por: Los Autores*

### 1.3.4.3. HISTOGRAMA

Al ser ésta una representación por áreas, hay que distinguir si los intervalos en los que aparecen agrupados los datos son de igual amplitud o no.[1]



*Gráfico I.3 Histograma*

*Fuente: Métodos Estadísticos Aplicados por Pedro Sánchez*

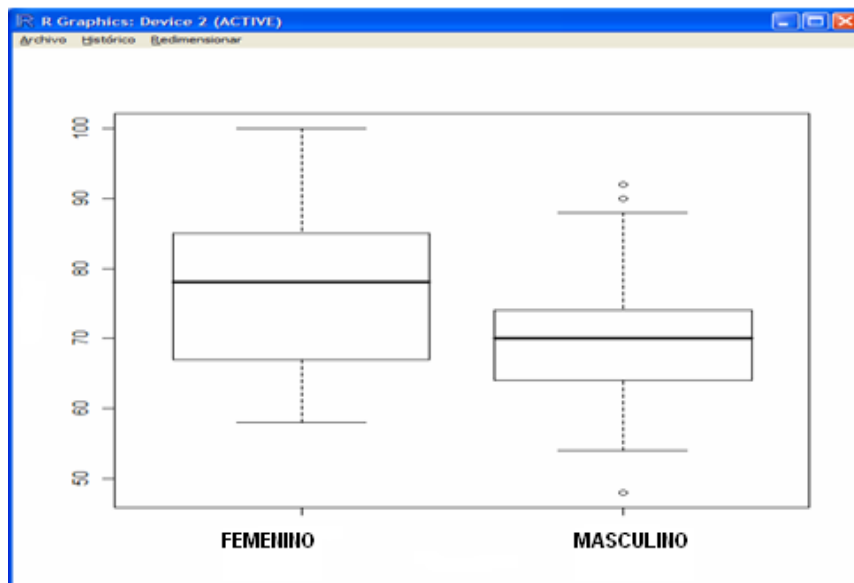
#### **1.3.4.4. DIAGRAMA DE CAJAS**

El diagrama de caja (boxplot) consiste en una caja cuyos bordes inferiores y superior son los cuartiles 1° y 3° y la línea central representa la mediana. Los bigotes desde la caja indican el rango de los datos. Además de elegir una variable cuantitativa, numérica, permite considerar un factor para comparar la primera respecto de los niveles del factor. [1]

**Cuartiles:** divide a la población o muestra en cuatro partes iguales.

**Deciles:** divide a la población en diez partes iguales.

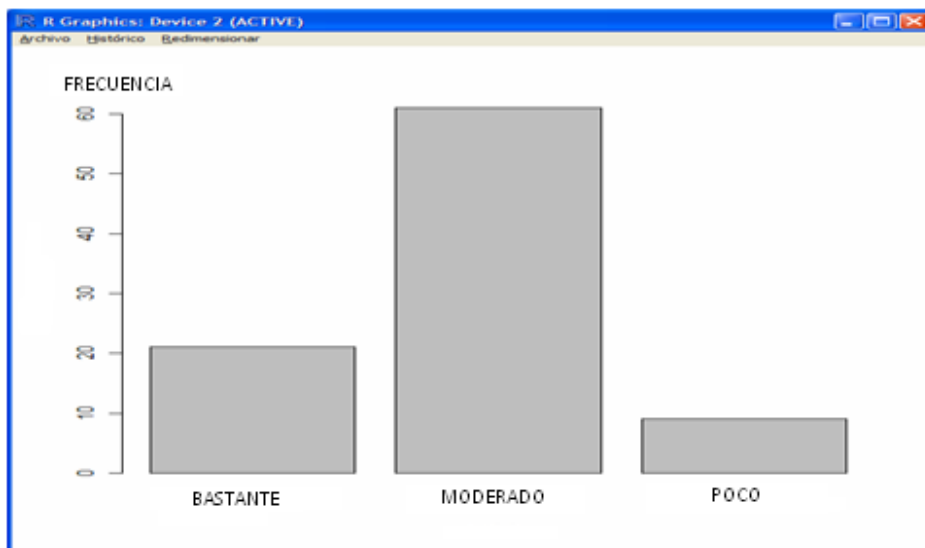
**Percentiles:** divide a la población en cien partes iguales.



*Gráfico I.4 Diagrama de cajas  
Fuente: Métodos Estadísticos Aplicados por Pedro Sánchez*

### 1.3.4.5. GRÁFICA DE BARRAS

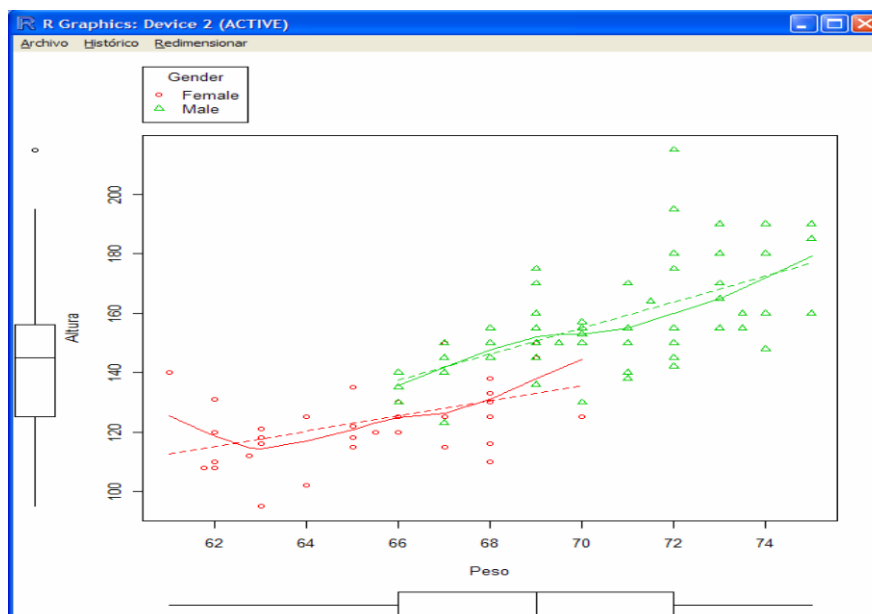
Gráfico usado para recoger las frecuencias de los niveles en las variables cualitativas (factores).[1]



*Gráfico I.5 Gráfica de barras  
Fuente: Métodos Estadísticos Aplicados por Pedro Sánchez*

### 1.3.4.6. DIAGRAMA DE DISPERSIÓN

Muestra conjuntamente datos de dos variables (en X y en Y) para ver su correlación, y permite considerar grupos (niveles de un factor).[3]

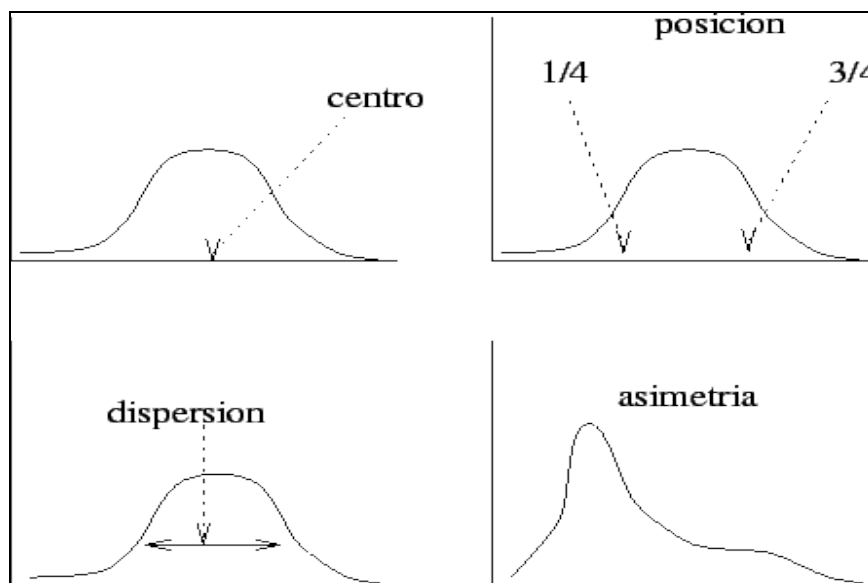


*Gráfico I.6 Diagrama de dispersión*

*Fuente: Medidas de dispersión y simetría de datos por Mario Suarez*

### 1.3.4.7. SIMETRÍA DE DATOS

Sabemos cómo calcular valores alrededor de los cuales se distribuyen las observaciones de una variable sobre una muestra y sabemos cómo calcular la dispersión que ofrecen los mismos con respecto al valor de central. Nos proponemos dar un paso más allá en el análisis de la variable. En primer lugar, nos vamos a plantear el saber si los datos se distribuyen de forma simétrica con respecto a un valor central, o si bien la gráfica que representa la distribución de frecuencias es *de una forma diferente del lado derecho que del lado izquierdo*. [3]



*Gráfico I.7 Medidas representativas de un conjunto de datos estadísticos  
Fuente: Medidas de dispersión y simetría de datos por Mario Suarez*

En este sentido pueden examinarse varias características, siendo las más comunes:

- ✓ La *tendencia central* de los datos;
- ✓ La *dispersión* o *variación* con respecto a este centro;
- ✓ Los datos que ocupan ciertas *posiciones*.
- ✓ La *simetría* de los datos.
- ✓ La *forma* en la que los datos se agrupan.

### **1.3.5. MEDIDAS DE TENDENCIA CENTRAL**

Los fenómenos biológicos no suelen ser constantes, por lo que será necesario que junto a una medida que indique el valor alrededor del cual se agrupan los datos, se asocie una medida que haga referencia a la variabilidad que refleje dicha fluctuación. Estudiaremos los estadísticos que nos van a orientar sobre cada uno de estos niveles de información: valores alrededor de los cuales se agrupa la muestra, la mayor o menor fluctuación alrededor de esos valores, nos interesaremos en ciertos valores que marcan posiciones características de una distribución de frecuencia.[4]

**Estadísticos de tendencia central:** las tres medidas más usuales de tendencia central son la *media, mediana y moda*.

### **1.3.5.1. MEDIA**

También llamada media aritmética, es una medida descriptiva que se calcula sumando los valores numéricos y dividiendo entre el número de valores. A la media poblacional se la conoce como  $\mu$  (mu) y la media muestral como  $\bar{x}$  (x barra). Cálculo de la media poblacional  $\mu$ :

$$\mu = \frac{\sum x}{N}$$

Donde  $N$  es el tamaño de la población y  $x$  son los valores de los datos de la población.

Cálculo de la media muestral  $\bar{x}$  :

$$\bar{x} = \frac{\sum x}{n}$$

Donde  $n$  es el tamaño de la muestra y  $x$  son todos los valores que toma la muestra.[4]

### **1.3.5.2. MEDIANA**

Es el valor central de la variable, es decir, supuesta la muestra ordenada en orden creciente o decreciente, el valor que divide en dos partes la muestra. La principal característica de esta medida es que al menos el 50% de las observaciones son menores o iguales a ella.

Cálculo de la mediana: cuando  $N$  es impar, hay un término central  $x_{\frac{N+1}{2}}$  que será el valor de la mediana. Cuando  $N$  es par, hay dos

términos centrales  $x_{\frac{N}{2}}, x_{\frac{N}{2}+1}$  la mediana será el promedio de estos dos valores.

### **1.3.5.3. MODA**

Es el valor de la variable que tenga mayor frecuencia absoluta, la que más se repite, es la única medida de centralización que tiene sentido estudiar en una variable cualitativa, pues no precisa la realización de cálculo alguno.[4]



### **1.3.6. MEDIDAS DE VARIABILIDAD O DISPERSIÓN**

Los estadísticos de *tendencia central* o *posición* nos indican donde se sitúa un grupo de puntuaciones. Los de *variabilidad* o *dispersión* nos indican si esas puntuaciones o valores están próximas entre sí o si por el contrario están o muy dispersas. Algunas de las principales medidas son:

#### **1.3.6.1 RANGO**

Se obtiene restando el valor más bajo de un conjunto de observaciones del valor más alto, se denota como  $R$ . Es fácil de calcular y sus unidades son las mismas que las de la variable, aunque posee varios inconvenientes:

- No utiliza todas las observaciones (sólo dos de ellas);
- Se puede ver muy afectada por alguna observación extrema;
- El rango aumenta con el número de observaciones, o bien se queda igual. En cualquier caso nunca disminuye.

#### **1.3.6.2 VARIANZA**

Es una medida de dispersión que mide la tendencia de las observaciones individuales a desviarse con respecto a la media. La varianza para los datos de una población está determinada por la siguiente ecuación:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Donde  $x$  son los valores de la población,  $\mu$  es la media poblacional y  $N$  es el número de observaciones en la población.[3]

#### **1.3.6.3. DESVIACIÓN ESTÁNDAR**

También mide la variabilidad de las observaciones con respecto a la media, es igual a la raíz cuadrada de la varianza. Esta medida de dispersión siempre es positiva y se denota por  $\sigma$ . Se calcula a través de la ecuación: [1]

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

### **1.3.7. NIVEL DE SIGNIFICANCIA**

Es la máxima cantidad de error que estamos dispuestos aceptar para dar como válida la hipótesis del investigador.

El investigador plantea una proposición y le asigna el valor de verdadero, al tomar tal decisión existe la probabilidad de equivocarse, de cometer un error, al cual denominamos error tipo I, entonces decide estimar la probabilidad de cometer ese error. El p-value definido como la probabilidad de que la proposición aceptada sea falsa no debe ser mayor al valor del error establecido convencionalmente en la fase de planificación como el límite de error que estamos dispuestos a aceptar.

**El error tipo I:** Ocurre cuando aceptamos la hipótesis del investigador, cuando tal proposición era falsa. Por lo tanto, es un juicio de valor equivocado.

**El error tipo II:** Ocurre cuando se acepta la hipótesis nula siendo esta falsa.

**El p-value:** Es la probabilidad de equivocarse al aceptar la hipótesis del investigador como verdadera; es decir la probabilidad de cometer un error tipo I.

**Variable Independiente:** Es aquella característica o propiedad que se supone ser la causa del fenómeno estudiado. En investigación experimental se llama así a la variable que el investigador manipula.

**Variable Dependiente:** Se define como propiedad o característica que se trata de cambiar mediante la manipulación de la variable independiente.

La variable dependiente es el factor que es observado y medido para determinar el efecto de la variable independiente.[2]

### **1.3.8. TIPOS DE ERRORES**

Los errores en las mediciones son de dos tipos: sistemáticos y aleatorios

**Errores sistemáticos.** Estos resultan de factores que comprenden el "sistema de medición" e incluyen el medio ambiente, los instrumentos y el observador. Siempre que las condiciones del sistema se mantengan constantes, los errores sistemáticos se mantendrán constantes. Si las condiciones cambian, las magnitudes de los errores sistemáticos también cambian.[2]

Las condiciones que ocasionan errores sistemáticos se deben a leyes físicas que se pueden representar matemáticamente. Por tanto, si se conocen las

condiciones y se pueden medir es posible calcular una corrección y aplicarla a los valores observados.

Un ejemplo de un error sistemático variable es el cambio de longitud de una cinta de acero como resultado de diferencias de temperatura que ocurren durante el tiempo de su utilización. Si se miden los cambios de temperatura, las correcciones de longitud se pueden calcular mediante una simple fórmula.

**Errores aleatorios.** Estos son los errores que quedan después de haber eliminado los errores sistemáticos. Son ocasionados por factores que quedan fuera del control del observador, obedecen las leyes de la probabilidad y se les llama también errores accidentales. Estos errores están presentes en todas las mediciones topográficas.[2]

Las magnitudes y los signos algebraicos de los errores aleatorios son consecuencia del azar. No existe una manera absoluta de calcularlos ni de eliminarlos, pero pueden estimarse usando un procedimiento de ajuste conocido como el método de mínimos cuadrados.

### **1.3.9. ANÁLISIS PARAMÉTRICO**

Las dójimas que con frecuencia se mencionan en los análisis siempre presuponen distribuciones particulares de la variable aleatoria, o bien sus hipótesis especifican parámetros o distribuciones. Por esto las estadísticas que se usan con más frecuencia son las estadísticas paramétricas.[1]

Estas técnicas permiten confirmar resultados o valorar las inconsistencias de ellos.

#### **¿Cuáles son los supuestos de las estadísticas paramétricas?**

1. La distribución poblacional de la variable dependiente es normal: el universo tiene una distribución normal.
2. El nivel de la medición de la variable dependiente es por intervalo o razón.
3. Cuando dos o más poblaciones son estudiadas, éstas tienen una varianza homogénea.
4. Las poblaciones en cuestión tienen una dispersión similar en sus distribuciones.

### 1.3.10. TIPOS DE ANOVA

#### 1.3.10.1 ANÁLISIS SIMPLE DE VARIANZA

El objetivo principal de muchos experimentos consiste en determinar el efecto que sobre alguna variable dependiente  $y$  tienen distintos niveles de algún factor  $x$  (variable independiente y discreta). El factor puede ser la temperatura, la empresa que ha producido el bien, el día de la semana, etc. Esencialmente, el diseño para el análisis simple de la varianza consistirá en obtener muestras aleatorias e independientes del valor de  $y$  asociado a cada uno de los distintos niveles del factor  $x_1, x_2, \dots, x_n$ . Entonces podremos determinar si los diferentes niveles del factor tienen un efecto significativo sobre el valor de la variable dependiente. El funcionamiento de la técnica ANOVA simple es, a grandes rasgos, el siguiente: a fin de comparar las medias de  $Y$  asociadas a los distintos niveles del factor  $(x_1, x_2, \dots, x_n)$ , compararemos una medida de la **variación entre diferentes niveles (MS-factor)** con una medida de la **variación dentro de cada nivel (MS-error)**. Si el MS-factor es significativamente mayor que el MS-error, concluiremos que las medias asociadas a diferentes niveles del factor son distintas. Esto significa que el factor influye significativamente sobre la variable dependiente  $y$ . Si, por el contrario, el MS-factor no es significativamente mayor que el MS-error, no rechazaremos la hipótesis nula de que todas las medias, asociadas a diferentes niveles del factor, coinciden.[7]

#### Supuestos

De forma similar a lo que ocurre con la regresión lineal, aquí también hay un modelo para los datos. El modelo asociado al  $i$ -ésimo nivel del factor  $x$  será:

$$Y = u_i + \varepsilon$$

Donde:

Los errores  $\varepsilon$  están normalmente distribuidos con media 0

Los errores  $\varepsilon$  son independientes

Los errores  $\varepsilon$  tienen varianza constante  $\sigma^2$

Para verificar estos supuestos suele ser útil realizar un gráfico que muestre la distribución de las observaciones por niveles: si en el gráfico se aprecian diferencias entre niveles por lo que a la variación de las observaciones se refiere, es muy probable que tengamos un problema con el supuesto de **varianza constante**; si aparecen “outliers”, puede que no se cumpla el supuesto de **normalidad**; por otra parte, si el tiempo fuese un factor importante a la hora de registrar observaciones, podría ocurrir que observaciones consecutivas estuviesen correlacionadas, con lo que no se cumpliría el supuesto de **independencia**.

Supóngase  $k$  grupos muestrales de igual tamaño  $p$ , siendo  $n$  el número de elementos muestrales ( $n = kp$ ), cada grupo procede de una población normal y todas tienen la misma varianza (desconocida).

Para cada valor muestral  $x_{ij}$  se le asignan **i j**, (i-ésimo grupo, j-ésimo valor). Los grupos muestrales tienen media  $\bar{x}$  y una varianza  $S^2$  y provienen de:

- variación dentro de la muestra (W)
- variación entre muestras (B).
- Media del grupo

$$\bar{x}_i = \frac{\sum_{j=1}^p x_{ij}}{p}$$

y

Media General

$$\bar{x} = \frac{\sum_{i=1}^k \sum_{j=1}^p x_{ij}}{n} = \frac{\sum_{i=1}^k p \bar{x}_i}{n} = \frac{\sum_{i=1}^k \bar{x}_i}{k}$$

<b>ANÁLISIS SIMPLE DE LA VARIANZA</b>			
<i>Origen de la varianza</i>	<i>Grados de libertad</i>	<i>Suma de diferencias cuadráticas (SDC)</i>	<i>Diferencias cuadráticas medias (DCM=SDC/v)</i>
<i>Entre grupos</i>	$v_1 = k - 1$	$SDC_B = p \sum_{i=1}^k (\bar{x}_i - \bar{x})^2$	$DMC_B = \frac{SDC_B}{k - 1}$
<i>Dentro del grupo</i>	$v_2 = n - k$	$SDC_w = \sum_{i=1}^k \sum_{j=1}^p (x_{ij} - \bar{x})^2$	$DMC_w = \frac{SDC_w}{n - k}$
<i>Total</i>	$v_T = n - 1$	$SDC_T = \sum_{i=1}^k \sum_{j=1}^p (x_{ij} - \bar{x})^2$	$DMC_T = \frac{SDC_T}{n - k}$

*Tabla I-3 Análisis simple de la varianza*  
*Fuente: Página de internet Autor Gustavo Alcalá*

Para aceptar que los grupos estudiados no son diferentes debe aplicarse la prueba F:

$$F_C = \frac{DMC_B}{DMC_w}$$

Comparar el valor de  $F_{cal}$  con  $F_{tabulado}$  para un valor de significación establecida  $\alpha$ , ( $v_1 = k-1$ ;  $v_2 = n-k$ )

Si  $F_{cal} < F_{tabulado}$  entonces no existen diferencias significativas

entre los distintos grupos muestrales.

En el caso de que existieran diferencias significativas entre los grupos al nivel  $\alpha$  preestablecido, se deberá considerar que al menos dos de las medias de grupos son diferentes entre sí.

### **1.3.10.2. ANÁLISIS DOBLE DE VARIANZA**

Usaremos el **análisis doble de varianza** para estudiar los posibles efectos causados por diferentes niveles de dos factores sobre la variable dependiente. Así, por ejemplo, en agricultura estaremos interesados en estudiar qué efectos tendrán, sobre el crecimiento de las papas, variaciones en los niveles de potasio y nitrógeno de la tierra; en medicina, estaremos interesados en estudiar los efectos, sobre el dolor de cabeza, del medicamento y de la dosis empleados; en educación, buscaremos conocer qué efectos, sobre el tiempo necesario para adquirir unos conocimientos, tendrán los factores nivel de estudios y género; en una campaña de marketing, estaremos interesados en conocer los efectos del presupuesto y del medio usado (televisión, revistas, ...) sobre las ventas; etc.[7]

Usaremos ANOVA doble para contrastar, para cada uno de los dos factores, la hipótesis nula de que el resultado de la variable dependiente (crecimiento de patatas, intensidad del dolor de cabeza, tiempo en adquirir conocimientos, ventas, etc.) no depende del factor.

#### **Modelo aditivo (sin interacción)**

El **modelo aditivo** supone que la variación total en los datos puede ser expresada como suma de variaciones procedentes de fuentes diversas:

(Variación total en los datos) = (Variación debida al primer factor) + (Variación debida al segundo factor) + (Variación debida al error aleatorio)

En el modelo anterior, si la variación debida al primer factor fuese mucho mayor que la variación debida al error aleatorio, dispondríamos de evidencia estadística contra la hipótesis nula de que los distintos niveles del primer factor tienen el mismo efecto sobre la variable dependiente; de forma similar, si la variación debida al segundo factor fuese mucho mayor que la variación debida al error aleatorio, deberíamos rechazar la hipótesis nula de que la variable dependiente no depende de los diversos niveles del segundo factor.

#### **Modelo con interacción**

Un aspecto al que debemos prestar especial atención es el **nivel de interacción entre ambos factores**; es decir, el efecto que cada uno de los factores tiene sobre el otro. Así, por ejemplo, en la campaña de marketing citada anteriormente, el incremento en las ventas debido a

un aumento del presupuesto podría ser el mismo independientemente del medio usado, o bien podría variar dependiendo del medio usado.

En este último caso deberíamos estudiar también la interacción entre los factores presupuesto y medio. Para ello usaremos los llamados **modelos con interacción**.

Este modelo supone que la variación total de los datos puede descomponerse de la siguiente forma: (Variación total en los datos) = (Variación debida al primer factor) + (Variación debida al segundo factor) + (Variación debida a la interacción entre factores) + (Variación debida al error aleatorio)

### **1.3.11. DISTRIBUCION "F" FISHER**

La necesidad de disponer de métodos estadísticos para comparar las varianzas de dos poblaciones es evidente a partir del análisis de una sola población. Frecuentemente se desea comparar la precisión de un instrumento de medición con la de otro, la estabilidad de un proceso de manufactura con la de otro o hasta la forma en que varía el procedimiento para calificar de un profesor universitario con la de otro.[10]

Intuitivamente, podríamos comparar las varianzas de dos poblaciones,  $\sigma_1^2$  y  $\sigma_2^2$ , utilizando la razón de las varianzas muestrales  $s_1^2/s_2^2$ . Si  $s_1^2/s_2^2$  es casi igual a 1, se tendrá poca evidencia para indicar que  $\sigma_1^2$  y  $\sigma_2^2$  no son iguales. Por otra parte, un valor muy grande o muy pequeño para  $s_1^2/s_2^2$ , proporcionará evidencia de una diferencia en las varianzas de las poblaciones.

La variable aleatoria F se define como el cociente de dos variables aleatorias ji-cuadradas independientes, cada una dividida entre sus respectivos grados de libertad. Esto es,

$$F = \frac{U/v_1}{V/v_2}$$



Donde  $U$  y  $V$  son variables aleatorias ji-cuadradas independientes con grados de libertad  $\nu_1$  y  $\nu_2$  respectivamente.

Sean  $U$  y  $V$  dos variables aleatorias independientes que tienen distribución ji cuadradas con  $\nu_1$  y  $\nu_2$  grados de libertad, respectivamente. Entonces la

distribución de la variable aleatoria  $F = \frac{U/\nu_1}{V/\nu_2}$  está dada por:

$$f(x) = \frac{\Gamma\left[\frac{(\nu_1 + \nu_2)}{2}\right] \left(\frac{\nu_1}{\nu_2}\right)^{n/2} x^{(n/2)-1}}{\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right) \left(1 + \nu_1 \frac{x}{\nu_2}\right)^{(\nu_1 + \nu_2)/2}}$$

$$0 < x < \infty$$

y se dice que sigue la distribución F con  $\nu_1$  grados de libertad en el numerador y  $\nu_2$  grados de libertad en el denominador.

La variable aleatoria F es no negativa, y la distribución tiene un sesgo hacia la derecha. La distribución F tiene una apariencia muy similar a la distribución ji-cuadrada; sin embargo, se encuentra centrada respecto a 1, y los dos parámetros  $\nu_1$  y  $\nu_2$  proporcionan una flexibilidad adicional con respecto a la forma de la distribución.

Si  $s_1^2$  y  $s_2^2$  son las varianzas muestrales independientes de tamaño  $n_1$  y  $n_2$  tomadas de poblaciones normales con varianzas  $\sigma_1^2$  y  $\sigma_2^2$ , respectivamente, entonces:

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \frac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2} = \left(\frac{s_1}{s_2}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right)^2$$

### **1.3.12. R CONSOLE (SOFTWARE)**

El software estadístico R aporta un entorno de trabajo orientado a resolver problemas de Estadística. A estos sistemas se les denomina a veces PSE, Problem Solving Environments, como Matlab, Mathematica, etc.). Es de uso libre en el dominio público, y resulta del esfuerzo cooperativo de personas e instituciones académicas relevantes relacionadas con la Estadística y la Computación en todo el mundo. Permite trabajar con una ventana de interacción con usuario, R Console, que ofrece posibilidades para gestionar archivos en disco, guardar resultados, etc.

Trabaja con un lenguaje de programación, y variables con diversos tipos (numéricas, cualitativas-factor, lógicas y caracter), y estructuras (escalar, vector, matriz, data.frame, lista), con constantes y con llamadas a funciones. El lenguaje utiliza elementos clásicos de programación, como funciones, instrucciones, expresiones aritméticas, lógicas, etc. Se puede trabajar tanto en modo programa, con un conjunto de instrucciones, como en modo de comandos interactivos. Ofrece un entorno básico interactivo a través de una ventana de diálogo.[12]

## **CAPÍTULO II**

### **2. ANÁLISIS DE LA INFORMACIÓN**

En el capítulo anterior hemos dado a conocer la forma operativa del proceso, por lo que el motivo de estudio de este capítulo será la descripción actual del proceso desde el punto de vista estadístico.

a) **Estandarización de proceso.**- Durante el año 2008 el proceso de elaboración de cubiertas galvanizadas permaneció en una fase de estandarización lo que implica establecimiento de manuales, instructivos de trabajo para el manejo de la máquina, formatos para registrar unidades producidas, el porcentaje de pérdida, la materia prima, chatarra e Índices de Productividad

b) **Recolección de datos.**-La organización ha establecido dentro de sus procedimientos llevar un registro por cada producción de cubierta galvanizada, para poder llevar la trazabilidad del producto a través de todo el proceso, iniciando desde los lotes de materia prima empleados hasta la obtención del producto final. Por tanto, contamos con una base de datos de un período de tres años 2009, 2010 y 2011.

c) **Muestreo de datos.**- El tamaño de muestra seleccionado, es  $N=984$  datos, al haber mayor cantidad de datos la muestra es más representativa y se reduce el error medio estándar.

d) **Tratamiento de los datos.**- El objetivo de tratar los datos es identificar el comportamiento de sus varianzas además de segregar los datos atípicos.

e) **Depuración de datos.**- En nuestra base datos es importante localizar los datos atípicos provenientes de mediciones que no se repiten durante el proceso y que por alguna razón puntual dentro del proceso, el porcentaje de pérdida se posiciona sobre o debajo de los demás datos recolectados. En estos casos se debe investigar la causa del surgimiento de estos datos atípicos y tratarlos con una acción correctiva o preventiva verificando tendencia.

Por medio de un análisis descriptivo podremos establecer los datos atípicos que se encuentran dentro de la base de datos  $N=984$

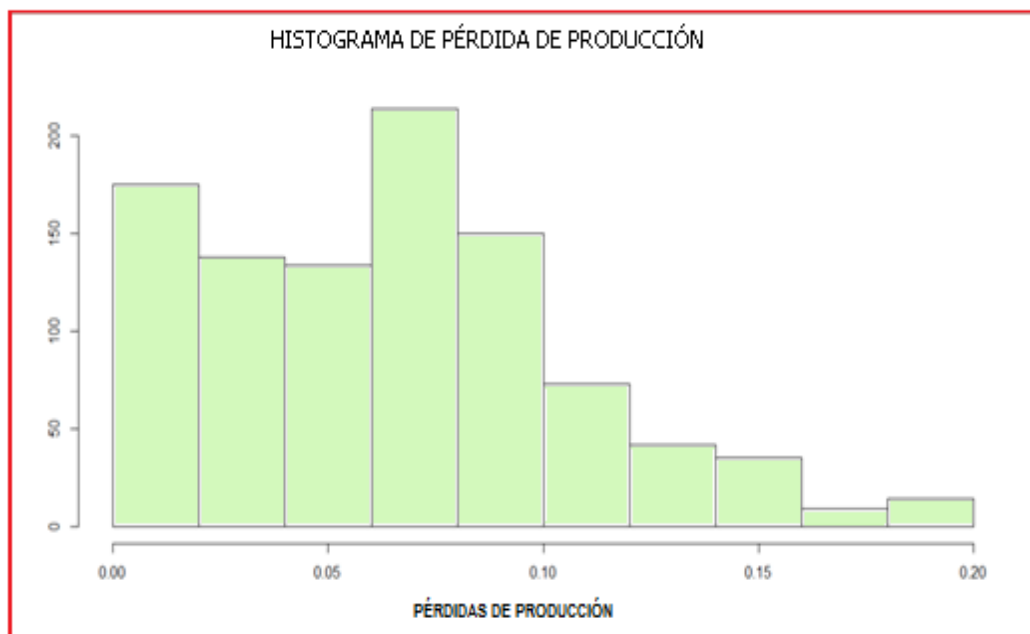
---

<b>N</b>	<b>984</b>
<b>X</b>	<b>6.4%</b>
<b>Sd</b>	<b>4.1%</b>
<b>Min</b>	<b>0%</b>
<b>Max</b>	<b>19.4%</b>
<b>Q1</b>	<b>2.9%</b>
<b>Q3</b>	<b>8.8%</b>

---

*Tabla II-1 Datos Paneladora#4  
Elaborado por: Los Autores*

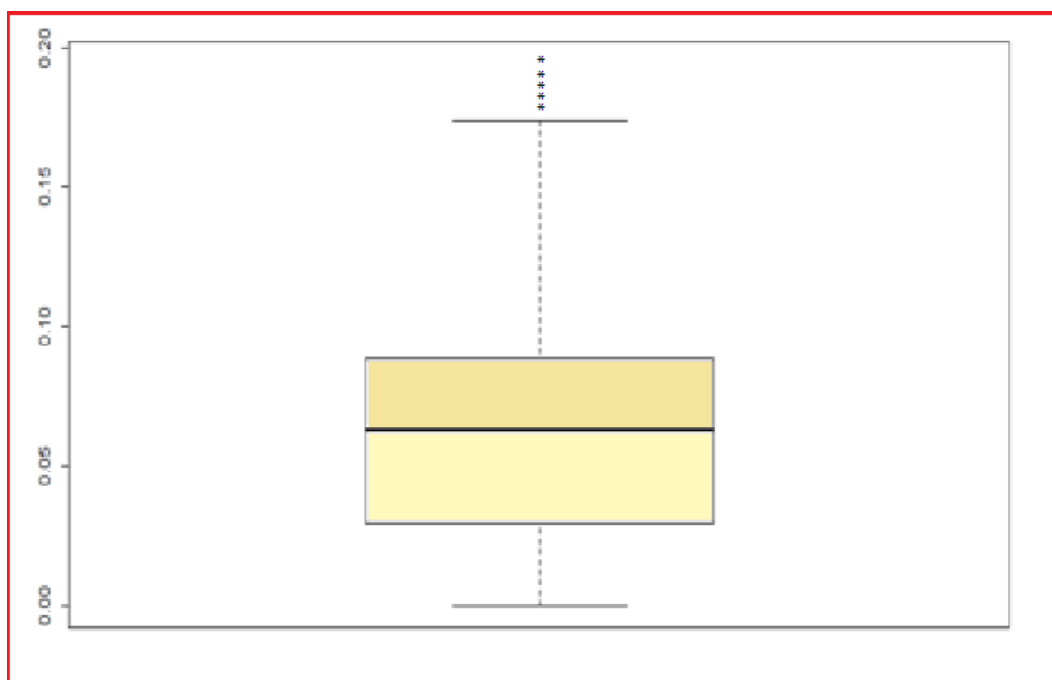
Los límites de porcentaje de pérdida que existe dentro del proceso va desde 0 a 19.4%, así como también podemos observar que la dispersión de los datos es de 4.1% con respecto al valor medio 6.4%



*Gráfico II.1 Histograma de pérdidas Paneladora#4  
Fuente: Estudio efectuado por los autores*

A partir de este histograma podemos concluir lo siguiente:

- 1.-El porcentaje de pérdida más bajo es 0 % y el más alto es de aproximadamente 19.4%
- 2.-La frecuencia de clase más grande va desde 6% a 8% de pérdida .Dentro de ese margen se encuentran más de 200 observaciones.
- 3.-La frecuencia de clase más pequeña va desde 16% a 18% de pérdida. Dentro de ese margen se encuentran aproximadamente 20 observaciones.



*Gráfico II.2 Diagrama de cajas de pérdidas Paneladora#4*

*Fuente: Estudio efectuado por los autores[11]*

El diagrama de caja muestra que en un 50% de los datos el porcentaje de pérdida se encuentra entre 2.9 y 8.8. La distancia entre los extremos de la caja, 5.9%, es el rango intercuartil. Este rango es la distancia entre el primer y el tercer cuartil; muestra la propagación o dispersión de la mayoría de los porcentajes de pérdida.

También revela que la distribución del porcentaje de pérdida tiene un sesgo positivo, en este caso hay dos fuentes de información que lo indican. Primero la línea punteada en la parte superior de la caja, que va desde 8.8 % (Q3) al valor máximo de 19.4%, es más larga que la línea punteada en la parte inferior de la caja que va de 2.9% (Q1) al valor mínimo de 0.00%. En otras palabras el 25% de los datos mayores que el tercer cuartil se encuentran más dispersos que el 25% de los datos menores en el primer cuartil.

Sobre aproximadamente el 17% de pérdida aparecen un conjunto de asteriscos (\*), un asterisco indica un dato atípico. Un dato atípico es un valor que no concuerda con el resto de los datos, el cual se determina con el siguiente cálculo:

**DATOS ATÍPICOS**

$>Q3+1.5(Q3-Q1)$	$> 17\%$
$< Q1-1.5(Q3-Q1)$	$<- 5\%$

*Tabla II-2 Datos atípicos Paneladora# 4  
Fuentes: Estudio efectuado por los autores*

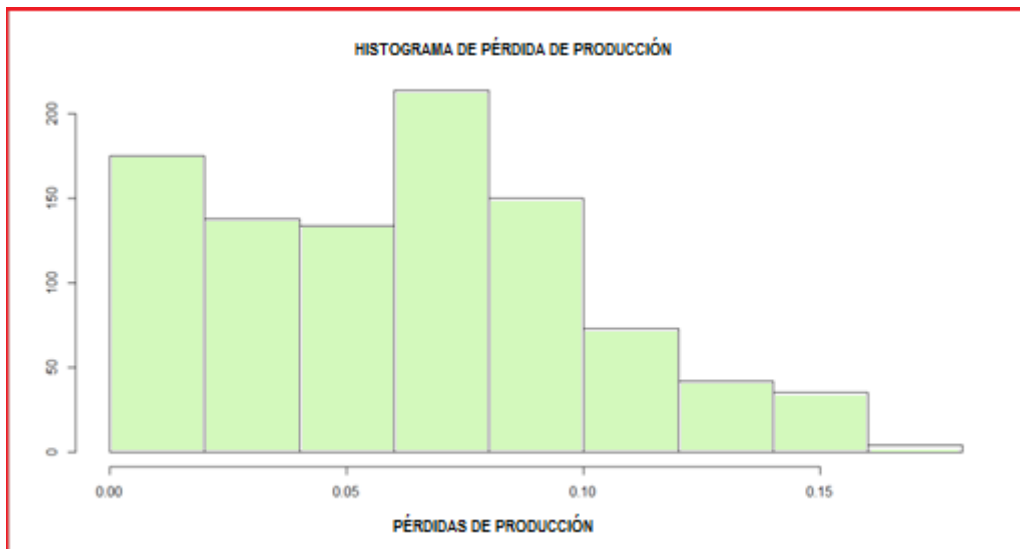
Procedemos a eliminar todas las observaciones cuyos valores sean mayores que el 17% completando un número de 19 observaciones eliminadas, para tener una base depurada que nos pueda ofrecer un estudio estadístico eficaz.

A continuación mostraremos el comportamiento de la base de datos  $N=965$  a través del diagrama de cajas y las medidas de tendencia central.

<b>N</b>	<b>965</b>
<b>X</b>	<b>6.2%</b>
<b>Sd</b>	<b>3.8%</b>
<b>Min</b>	<b>0.0%</b>
<b>Max</b>	<b>16.5%</b>
<b>Q1</b>	<b>2.9%</b>
<b>Q3</b>	<b>8.7%</b>

*Tabla II-3 Datos depurados Paneladora# 4  
Fuente: Estudio efectuado por los autores*

Los límites de porcentaje de pérdida que existe dentro del proceso va desde 0 % a 16.5%, así como también podemos observar que la dispersión de los datos es de 3.8% con respecto al valor medio 6.2%

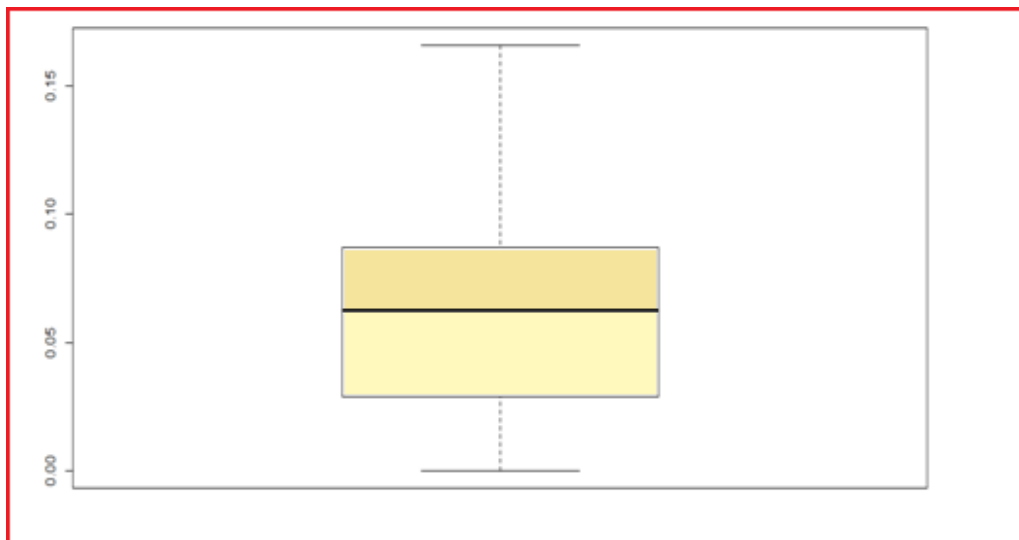


*Tabla II-4 Histograma: datos depurados de pérdidas Paneladora# 4*

*Fuente: Estudio efectuado por los autores*

- 1.-El porcentaje de pérdida más bajo es 0 % y el más alto es de aproximadamente 16.5 %
- 2.-La frecuencia de clase más grande va de 6 % a 8 % de pérdida. Dentro de ese margen se encuentran más de 200 observaciones.

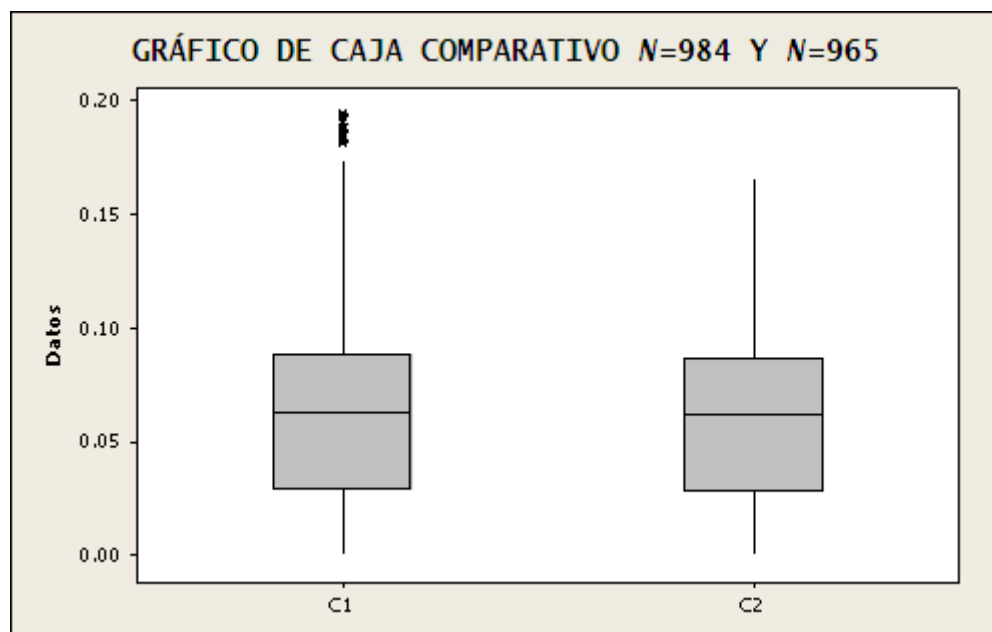




*Gráfico II.3 Diagrama de cajas: datos depurados de pérdidas Paneladora# 4  
Fuente: Estudio efectuado por los autores*

El diagrama de caja muestra que en un 50% de los datos el porcentaje de pérdida se encuentra entre 2.9 % y 8.7 %. La distancia entre los extremos de la caja, 5.8 % es el rango intercuartil.

También revela que la distribución del porcentaje de pérdida tiene un sesgo positivo, en este caso hay dos piezas de información que lo sugieren. Primero la línea punteada en la parte superior de la caja, que va desde 8.7 %(Q3) al valor máximo de 16.5%, es más larga que la línea punteada en la parte de la caja que va de 2.7%(Q1) al valor mínimo de 0.00%. En otras palabras el 25% de los datos mayores que el tercer cuartil se encuentran más dispersos que el 25% menor que el primer cuartil.



*Gráfico II.4 Diagrama de cajas: comparativo de pérdidas Paneladora# 4  
Fuente: Estudio efectuado por los autores*

Al eliminar los datos atípicos en el primer grupo donde  $N=984$  observaciones, se pudo constatar que la desviación estándar varió de 4.1% a 3.8 % en función al nuevo grupo con  $N= 965$  observaciones, podemos deducir que después de eliminar los datos atípicos las observaciones disminuyeron su dispersión con respecto al valor medio.

## **2.1.DISEÑO EXPERIMENTAL**

### **Objetivo:**

Determinar la variación del porcentaje de pérdida, por influencia de variables tales como los turnos u operarios.

Una vez establecido los datos que se someterán a estudio, se considerará como función respuesta al porcentaje de pérdida y como posibles factores de influencia las variables turno y operarios. Se ha recabado información histórica del proceso de producción de cubiertas galvanizadas en los tres diferentes turnos mañana, tarde y noche, con la participación de cinco operarios.

**Herramientas Estadísticas:**

Formulación de hipótesis:

ANOVA one way: 1 factor

ANOVA two way: 2 factores

Distribución de F de Fisher como parte del contraste de hipótesis.

Por motivos de explicación asumimos que los supuestos del modelo se cumplen:

1.-Siguen distribución normal.

2.-Las varianzas son homogéneas.

Continuando con el análisis en el programa R-Studio, se ha dispuesto los siguientes modelos:

**Modelo 1.-** Análisis de Influencia por la variable Turno

**Modelo 2.-** Análisis de Influencia por la variable Operario

**Modelo 3.-** Análisis de variables sin interacción

**Modelo 4.-** Análisis de variables con interacción

## 2.2. DESARROLLO

### 2.2.1. MODELO 1: ANÁLISIS DE INFLUENCIA POR VARIABLE TURNO.

Este modelo trata de explicar la influencia de la variable turno en el porcentaje de pérdidas

**Hipótesis:**

$$H_0 = u_1 = u_2 = u_3$$

$$H_1 = u_1 \neq u_2 \neq u_3$$

Donde:

**$H_0$ :** Las medias de los tratamientos son iguales.

**$H_1$ :** Por lo menos una de las medias de los tratamientos es diferente.

RESUMEN				
Grupos	Cuenta	Suma	Promedio	Varianza
MAÑANA	473	31,23685377	0,06603986	0,00148529
TARDE	227	12,566734	0,055360062	0,00136368
NOCHE	265	16,4040369	0,061902026	0,00145626

*Tabla II-5 Análisis de varianza de un factor (pérdidas –turno)*

*Fuente: Estudio efectuado por los autores*

<b>ANÁLISIS DE VARIANZA</b>		
<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>
Entre grupos	0,0176	2,0000
Dentro de los grupos	1,3937	962,0000
Total	1,4113	964,0000
<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
6,0682	0,0024	3,0051

*Tabla II-6 Análisis de varianza de un factor (pérdidas –turno)  
Fuente: Estudio efectuado por los autores*

Normalmente la variación entre grupos debería ser mayor que la variación dentro de los grupos, lo cual podemos identificarlo a través de los promedios de los cuadrados.

$$DCM_E > DCM_W$$

La variación entre grupos nos indica la variación que existe en el porcentaje de pérdida entre los tres turnos mañana, tarde y noche siendo el promedio de los cuadrados  $DCM_E = 0.00879$  y la variación dentro de las observaciones de un solo turno, sea este mañana, tarde o noche está dada por el promedio de los cuadrados  $DCM_W = 0.00144$ .

Ante esto podemos concluir que se cumple la condición lógica:

$$0.00879 > 0.00144$$

<b>Call:</b>				
lm(fórmula = PÉRDIDA ~ TURNO, data = Base. pérdida)				
<b>Residuals:</b>				
<b>Min</b>	<b>1Q</b>	<b>Median</b>	<b>3Q</b>	<b>Max</b>
-0.066040	-0.033465	0.000446	0.023489	0.110461
<b>Coefficients:</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
(Intercept)	0.066040	0.001750	37.735	< 2e-16 ***
<b>TURNONOCHE</b>	-0.004138	0.002921	-1.417	0.156872
<b>TURNOTARDE</b>	-0.010680	0.003073	-3.475	0.000534 ***
<b>Signif. codes:</b> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
<b>Residual standard error:</b> 0.03806 on 962 degrees of freedom				
<b>Multiple R-squared:</b> 0.01246, <b>Adjusted R-squared:</b> 0.01041				
<b>F-statistic:</b> 6.068 on 2 and 962 DF, <b>p-value:</b> 0.002405				

*Tabla II-7 Análisis de medias entre turnos (pérdidas Paneladora #4)  
Fuente: Estudio efectuado por los autores*

Para un mejor entendimiento a nuestro análisis nombraremos como intercepto (intercept) al Turno Mañana.

Podemos decir que se estima la media con un error bastante pequeño, como podemos observar a continuación el error estándar de la media de mañana es de 0.001750, la diferencia del turno de la noche se estima con un error de 0.00292 y la diferencia del turno de la tarde se estima con un error de 0.00307.

El error de estimación es bastante pequeño porque hay una gran cantidad de datos que nos permite tener una muestra representativa.

Observamos que las diferencias entre el turno de la noche con el turno de la mañana es de -0.0041, lo cual nos señala que no existe mucha diferencia significativa entre la pérdida producida en la mañana y en la noche. Sin

embargo la diferencia entre el turno de la tarde con el turno de la mañana es de -0.0106, lo cual indica que existe una diferencia significativa entre la pérdida producida en la mañana y en la tarde.

TURNO	MAÑANA	TARDE	NOCHE
MEDIA	6.603%	5.536%	6.1902%

*Tabla II-8 Tabla de medias por turno (pérdidas Paneladora 4)  
Fuente: Estudio efectuado por los autores*

Según la tabla de las medias por turno se puede identificar que en el turno de la tarde se produce menos pérdida comparado con los turnos de la mañana y noche.

Por lo tanto a través del estadístico p-value podemos aceptar la  $H_1$ : *Por lo menos una de las medias de los tratamientos es diferente.*

A continuación haremos una comparación entre la variabilidad total del proceso y la variabilidad influenciada por los turnos, a través de ANOVA y planteamos la siguiente hipótesis:

$H_0$ : *La variable turno influye significativamente en la función respuesta pérdida.*

$H_1$ : *La variable turno influye significativamente en la función respuesta pérdida.*

Analysis of Variance Table						
Model 0: PÉRDIDA ~ 1						
Model 1: PÉRDIDA ~ TURNO						
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	964		1.4113			
2	962	1.3937	2	0.017583	6.0682	0.002405 **
---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Tabla II-9 Comparación entre  $RSS_0$  y  $RSS_1$  (pérdidas Paneladora 4)  
Fuente: Estudio efectuado por los autores

El  $RSS_0$  es mayor que  $RSS_1$  debido a que se recoge toda la variabilidad del proceso, pero en el modelo 1 (Model 1) sólo se observa la variabilidad del porcentaje de pérdida con respecto a los turnos

$$RSS_0 < RSS_1$$

### REGLAS DE DECISIÓN:

Podemos basar nuestra decisión con cualquiera de los dos estadísticos de prueba que mencionaremos a continuación:

#### F-Fisher

- Si  $F_{cal} < F_{tabulado}$  se acepta la hipótesis nula.
- Si  $F_{cal} > F_{tabulado}$  se acepta la hipótesis alterna.

El  $F_{cal} = 6.0682 > F_{tab} = 3.0050$  entonces se acepta  $H_1$ : La variable turno influye significativamente en la función respuesta pérdida.

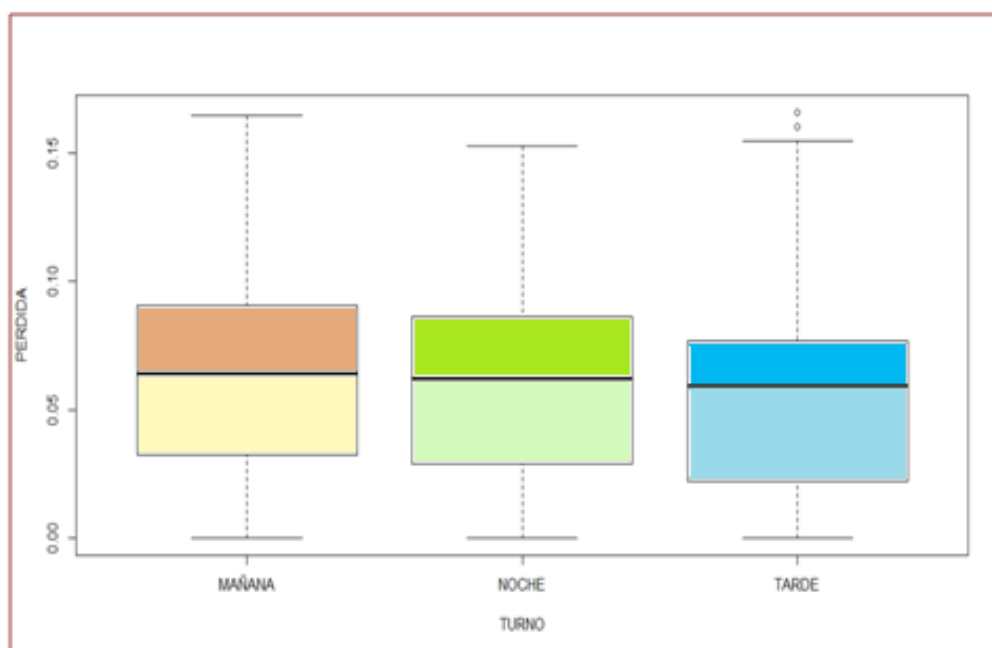


### **p-Value**

Si el p-value es mayor que el nivel de significancia ( $\alpha = 0.05$ ), se acepta la hipótesis nula

Si el p-value es menor que el nivel de significancia ( $\alpha = 0.05$ ), se acepta la hipótesis alterna

El p-value 0.0024 es menor que el nivel de significancia entonces *se acepta la  $H_1$ : La variable turno influye significativamente en la función respuesta pérdida.*



*Gráfico II.5 Diagrama de cajas: medias comparativas de turnos Paneladora 4  
Fuente: Estudio efectuado por los autores*

En el turno de la mañana es cuando se produce mayor pérdida, en la noche no hay mucha variación con respecto a la mañana, pero en la tarde es el mejor turno debido a que se produce menos pérdida durante la producción.

### 2.2.2. MODELO 2: ANALISIS DE INFLUENCIA POR VARIABLE OPERARIO

Este modelo explica la influencia de la variable operador en el porcentaje de pérdidas.

**Hipótesis:**

$$H_0 = u_1 = u_2 = u_3 = u_4 = u_5$$

$$H_1 = u_1 \neq u_2 \neq u_3 \neq u_4 \neq u_5$$

**Donde:**

**$H_0$ :** Las medias de los tratamientos son iguales.

**$H_1$ :** Por lo menos una de las medias de los tratamientos es diferente.

RESUMEN				
Grupos	Cuenta	Suma	Promedio	Varianza
Operario 1	206	10,7963	0,0524	0,0012
Operario 2	376	25,9615	0,0690	0,0015
Operario 3	33	2,2288	0,0675	0,0011
Operario 4	31	2,2693	0,0732	0,0015
Operario 5	319	18,9517	0,0594	0,0015

*Tabla II-10 Análisis de varianza de un solo factor (pérdidas-operario)  
Fuente: Estudio efectuado por los autores*

<b>ANÁLISIS DE VARIANZA</b>		
<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>
Entre grupos	0,0445	4,0000
Dentro de los grupos	1,3668	960,0000
Total	1,4113	964,0000
<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
7,8165	0,0000034	2,3812

*Tabla II-11 Análisis de varianza de un solo factor (pérdidas-operario)  
Fuente: Estudio efectuado por los autores*

La variación entre grupos nos indica la variación que existe en el porcentaje de pérdida entre los cinco operarios siendo el promedio de los cuadrados  $DCM_E = 0.01112$  y la variación dentro de las observaciones de un solo operario, está dada por el promedio de los cuadrados  $DCM_W = 0.00142$ .

Ante esto se concluye que se cumple la condición lógica:

$$0.01112 > 0.00142$$

<b>Call:</b>				
lm(fórmula = PÉRDIDA ~ OPERADOR, data = Base. pérdida)				
<b>Residuals:</b>				
<b>Min</b>	<b>1Q</b>	<b>Median</b>	<b>3Q</b>	<b>Max</b>
-0.069047	-0.034312	0.001833	0.022886	0.113411
<b>Coefficients:</b>				
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value Pr(&gt; t )</b>
(Intercept)		0.052409	0.002629	19.936 < 2e-16 ***
OPERADOR 2		0.016637	0.003271	5.087 4.38e-07 ***
OPERADOR 3		0.015129	0.007075	2.138 0.03274 *
OPERADOR 4		0.020794	0.007269	2.861 0.00432 **
OPERADOR 5		0.007000	0.003373	2.076 0.03819 *
<b>-Signif. codes:</b> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
<b>Residual standard error:</b> 0.03773 on 960 degrees of freedom				
<b>Multiple R-squared:</b> 0.03154, <b>Adjusted R-squared:</b> 0.02751				
<b>F-statistic:</b> 7.816 on 4 and 960 DF, <b>p-value:</b> 3.364e-06				

*Tabla II-12 Análisis de medias entre operarios (pérdidas Paneladora 4)*  
*Fuente: Estudio efectuado por los autores*

Para un mejor entendimiento a nuestro análisis nombraremos como intercepto (intecept) al operador 1

Podemos decir que se estima la media con un error bastante pequeño, como podemos observar a continuación el error estándar de la media del Operador 1 es 0.002629, la diferencia promedio del Operador 2 se estima con un error de 0.003271, la diferencia promedio del Operador 3 se estima con un error de 0.007075, la diferencia promedio del Operador 4 se estima con un error de 0.007269, la diferencia promedio del Operador 5 se estima con un error de 0.003373. El error de estimación es bastante pequeño porque hay una gran cantidad de datos que nos permite tener una muestra representativa.

Observamos que las diferencias entre la media de pérdida del Operador 1 siendo este el intercepto 0.052409, con respecto a el Operador 3 y Operador 5 no es muy significativa sin embargo las medias de pérdidas producidas por el operador 2 son bastante significativas comparadas con la media del operador 1 así mismo podemos observar que la media producida por el operador 2 es medianamente significativa.

OPERARIOS	MEDIA
OPERADOR 1	5.240%
OPERADOR 2	6.904%
OPERADOR 3	6.753%
OPERADOR 4	7.320%
OPERADOR 5	5.940%

*Tabla II-13 Medias de pérdidas por operarios (pérdidas Paneladora 4)*

*Fuente: Estudio efectuado por los autores*

Según la tabla de las medias por operario se observa que el operario 1 produce menos pérdida que todos los demás operarios; los operarios Jaime 2 y 4 producen más pérdida que los operarios 3 y 5

Por lo tanto a través del estadístico p-value podemos aceptar la  **$H_1$** : *Por lo menos una de las medias de los tratamientos es diferente.*

A continuación haremos una comparación entre la variabilidad total del proceso y la variabilidad influenciada por los operarios, a través de ANOVA y planteamos la siguiente hipótesis:

**$H_0$ :** *La variable operario no influye significativamente en la función respuesta pérdida.*

**$H_1$ :** *La variable operario influye significativamente en la función respuesta pérdida.*

Analysis of Variance Table						
Model 0: PÉRDIDA ~ 1						
Model 2: PÉRDIDA ~ OPERADOR						
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	964	1.4113				
2	960	1.3668	4	0.044514	7.8165	3.364e-06 ***
---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

*Tabla II-14 Comparación entre RSS0 y RSS2 (pérdidas Paneladora 4)  
Fuente: Estudio efectuado por los autores*

El  $RSS_0$  es mayor que  $RSS_2$  debido a que se recoge toda la variabilidad del proceso en el modelo cero, pero en el modelo 2 permite a cada operador tener su propia media y por ende el nivel de error es menor.

Sólo se observa la variabilidad del porcentaje de pérdida versus operarios.

$$RSS_0 < RSS_2$$

### **REGLAS DE DECISIÓN:**

Podemos basar nuestra decisión con cualquiera de los dos estadísticos de prueba que mencionaremos a continuación:

#### **F-Fisher**

- Si  $F_{cal} < F_{tabulado}$  se acepta la hipótesis nula.
- Si  $F_{cal} > F_{tabulado}$  se acepta la hipótesis alterna.

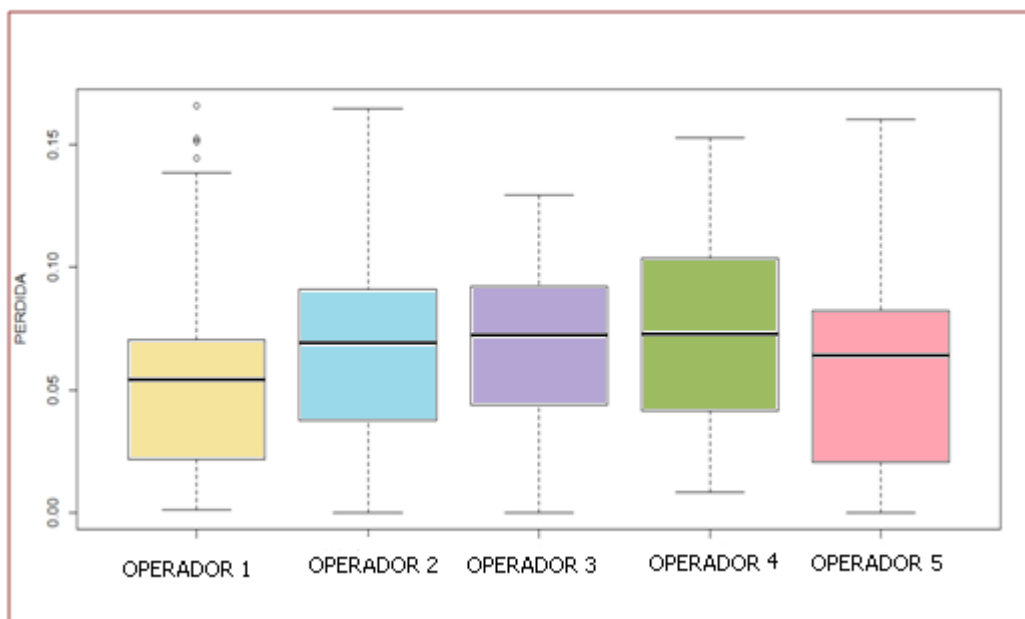
El  $F_{cal} = 7,81647091 > F_{tabulado} = 3,3637E-06$ , entonces **se acepta la  $H_1$ : La variable operario influye significativamente en la función respuesta pérdida.**

#### **p-Value**

Si el p-value es mayor que el nivel de significancia ( $\alpha = 0.05$ ), se acepta la hipótesis nula

Si el p-value es menor que el nivel de significancia ( $\alpha = 0.05$ ), se acepta la hipótesis alterna

El p-value 0.000003 es menor que el nivel de significancia entonces **se acepta la  $H_1$ : La variable operario influye significativamente en la función respuesta pérdida.**



*Gráfico II.6 Diagrama de cajas: medias comparativas de operarios (Paneladora 4)  
Fuente: Estudio efectuado por los autores*

El operador 5 produce menos pérdida después del operador 1 y el operador 4 sólo tiene 31 observaciones por lo que no hay suficiente evidencia para concluir que produce mayor pérdida, lo contrario del operador 2 con 376 observaciones podemos señalar que existe suficiente evidencia para concluir que produce mayor pérdida.

### **2.2.3 MODELO 3: ANÁLISIS DE VARIABLES SIN INTERACCIÓN**

En este modelo se hace el supuesto que la media de la diferencia entre los turnos es la misma para todos los operadores o equivalentemente que la diferencia entre operadores es la misma en todos los turnos.

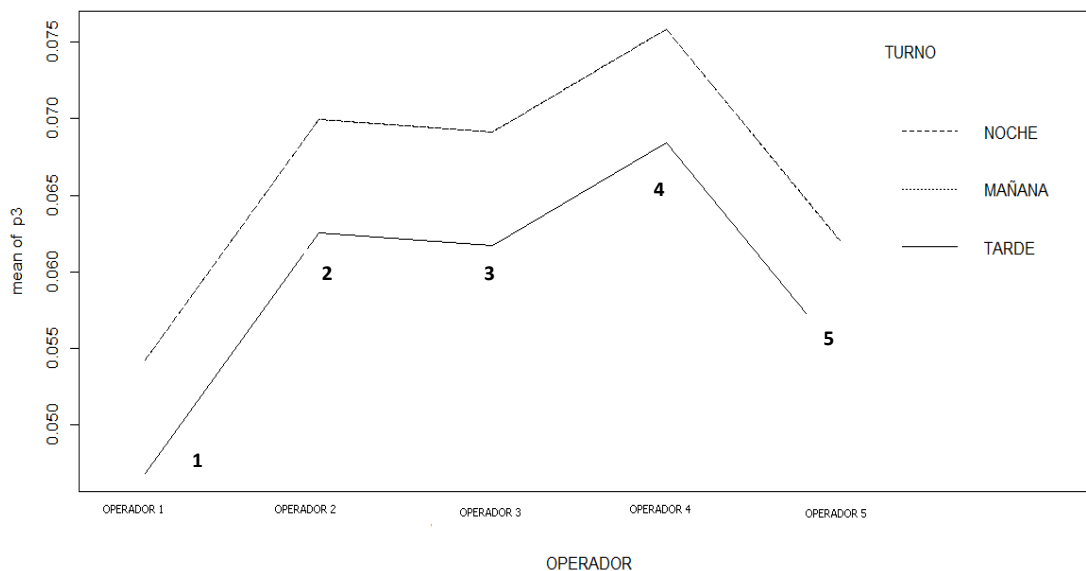


Call:				
lm(fórmula = PÉRDIDA ~ TURNO + OPERADOR, data = Base. pérdida)				
Residuals:				
Min	1Q	Median	3Q	Max
-0.069974	-0.033995	0.001691	0.022494	0.118987
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.422e-02	3.283e-03	16.514	< 2e-16 ***
TURNONOCH	2.924e-05	3.037e-03	0.010	0.99232
TURNOTARDE	-	3.161e-03	-2.338	0.01960 *
OPERADOR 2	1.572e-02	3.445e-03	4.564	5.68e-06 ***
OPERADOR 3	1.488e-02	7.136e-03	2.085	0.03735 *
OPERADOR 4	2.159e-02	7.297e-03	2.959	0.00316 **
OPERADOR 5	7.774e-03	3.472e-03	2.239	0.02537 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 0.03765 on 958 degrees of freedom				
Multiple R-squared: 0.03791, Adjusted R-squared: 0.03188				
F-statistic: 6.291 on 6 and 958 DF, p-value: 1.683e-06				

*Tabla II-15 Análisis sin interacción entre variables turno, pérdida y operario  
Fuente: Estudio efectuado por los autores*

Para un mejor entendimiento a nuestro análisis nombraremos como intercepto (intecept) al operador 1 en el turno Mañana.

Se representa el modelo sin interacción es decir que la diferencia entre las medias de todos los operarios es iguales en los tres turnos y viceversa.



**Gráfico II.7 Pronóstico sin Interacción entre turnos y operarios (Paneladora 4)**  
**Fuente: Estudio efectuado por los autores**

En el gráfico podemos identificar con un punto rojo las medias de pérdida producidas por operarios en los tres turnos, podemos visualizar también que en el turno de la mañana y noche se produce más pérdida que en la tarde asumiendo que no existe interacción entre variables turno y operador. Adicionalmente vamos a enlistar a cada operario según sus promedios de pérdidas producidos, en orden ascendente.

Operador 1, Operador 5, Operador 3, Operador 2, Operador 4.

## 2.2.4 MODELO 4: ANÁLISIS DE VARIABLES CON INTERACCIÓN

En este modelo se analizará el efecto de la variable turno sobre la función respuesta pérdida que difiere según el valor de la otra variable operador

### Hipótesis:

$H_0$ : La interacción no es significativa

$H_1$ : La interacción es significativa

Analysis of Variance Table					
Model 3: PÉRDIDA ~ TURNO + OPERADOR					
Model 4: PÉRDIDA ~ TURNO + OPERADOR + TURNO:OPERADOR					
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	919		1.0660		
2	911	1.0422	8	0.02377	2.5972 0.008254 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Tabla II-16 Comparación entre modelo aditivo y modelo con interacción (pérdidas Paneladora 4)  
Fuente: Estudio efectuado por los autores

### REGLAS DE DECISIÓN:

#### p-Value

Si el p-value es mayor que el nivel de significancia ( $\alpha = 0.05$ ), se acepta la hipótesis nula

Si el p-value es menor que el nivel de significancia ( $\alpha = 0.05$ ), se acepta la hipótesis alterna

El p-value 0.008254 es menor que el nivel de significancia entonces se acepta  $H_1$ : La interacción es significativa.

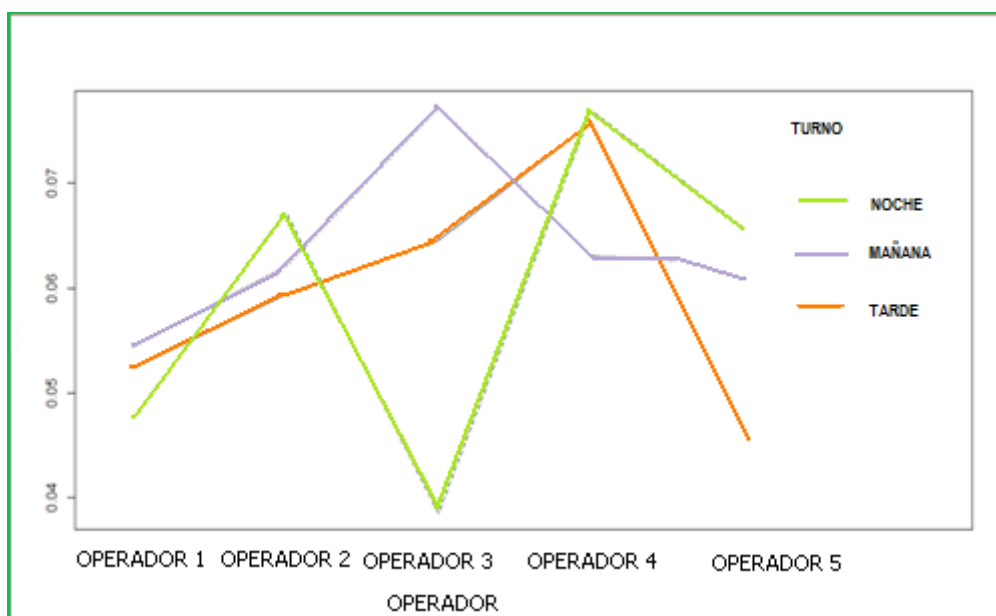
<b>Call:</b>					
lm(fórmula = PÉRDIDA ~ TURNO + OPERADOR + TURNO:OPERADOR, data = Base. pérdida)					
<b>Residuals:</b>					
<b>Min</b>	<b>1Q</b>	<b>Median</b>	<b>3Q</b>	<b>Max</b>	
-0.079575	-0.033073	0.001583	0.023911	0.111956	
<b>Coefficients:</b>		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
(Intercept)		0.0562286	0.0052956	10.618	<2e-16 ***
TURNONOCH		-0.0066924	0.0064340	-1.040	0.2985
TURNOTARDE		-0.0016483	0.0074522	-0.221	0.8250
OPERADOR 2		0.0121450	0.0058029	2.093	0.0366 *
OPERADOR 3		0.0210286	0.0099071	2.123	0.0340 *
OPERADOR 4		0.0068975	0.0116576	0.592	0.5542
OPERADOR 5		0.0078468	0.0061634	1.273	0.2033
TURNONOCH:OPERADOR 2		0.0106123	0.0080211	1.323	0.1861
TURNOTARDE:OPERADOR 2		0.0002474	0.0095733	0.026	0.9794
TURNONOCH:OPERADOR 3		-0.0320239	0.0185794	-1.724	0.0851.
TURNOTARDE:OPERADOR 3		-0.0109837	0.0180541	-0.608	0.5431
TURNONOCH:OPERADOR 4		0.0313775	0.0186965	1.678	0.0936.
TURNOTARDE:OPERADOR 4		0.0143383	0.0170546	0.841	0.4007
TURNONOCH:OPERADOR 5		0.0108146	0.0085197	1.269	0.2046
TURNOTARDE:OPERADOR 5		-0.0140695	0.0088317	-1.593	0.1115
<b>Signif. codes:</b> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
<b>Residual standard error:</b> 0.03745 on 950 degrees of freedom					
<b>Multiple R-squared:</b> 0.05615, <b>Adjusted R-squared:</b> 0.04224					
<b>F-statistic:</b> 4.037 on 14 and 950 DF, <b>p-value:</b> 7.861e-07					

**Tabla II-17 Análisis de la interacción entre variables turno, pérdidas y operario**  
*Fuente: Estudio efectuado por los autores*

Para un mejor entendimiento a nuestro análisis nombraremos como intercepto (intecept) al Operador 1 en el turno Mañana.

TURNO	OPERARIOS				
	OPERADOR 1	OPERADOR 2	OPERADOR 3	OPERADOR 4	OPERADOR 5
MAÑANA	5,623%	6,837%	7,726%	6,313%	6,408%
NOCHE	4,954%	6,684%	2,420%	8,761%	6,704%
TARDE	5,458%	5,648%	4,524%	7,057%	4,216%

*Tabla II-18 Porcentaje de pérdida para cada operario por turno (pérdidas Paneladora 4)  
Fuente: Estudio efectuado por los autores*



*Gráfico II.8 Pronósticos con interacción entre turnos y operarios (Paneladora 4)  
Fuente: Estudio efectuado por los autores*

Una manera de estudiar la interacción es al graficar medias de factores en una gráfica denominada de interacción. Para completar el estudio se debe explorar la interacción entre los turnos y los operarios. El trazo de la gráfica inicia con la colocación de los puntos que representan las medias de pérdidas.

Con esta gráfica se comprende mejor la interacción entre los efectos de los operarios y los turnos sobre el porcentaje de pérdida. Si los segmentos de recta fueran paralelos tal vez no haya interacción. Por otro lado, si los segmentos de recta se cruzan, esto sugiere una interacción entre las variables.

Además podemos concluir lo siguiente:

- Se evidencia que el operador 1 tiene un menor porcentaje de pérdidas en comparación a los operadores: Operador 2, Operador 4 y Operador 5 en los turnos: mañana, tarde y noche. Y en comparación al operador 3 tiene menor pérdida solo en el turno de la mañana.
- El operador 2 tiene un mayor porcentaje de pérdida que el operador 1 en los turnos: mañana, tarde y noche y mayor porcentaje de pérdida que el operador 3 en los turnos: tarde y noche, pero en comparación a los operadores 4 y 5 su porcentaje de pérdida es menor en los tres turnos.
- El operador 3 tiene mayor porcentaje de pérdida en el turno de la mañana en comparación con los operadores 1,2,4 y 5, pero muestra el menor porcentaje de pérdidas entre el grupo de operadores en los turnos: tarde y noche.
- El operador 4 tiene menor porcentaje de pérdida trabajando en la mañana en comparación a los turnos: tarde y noche.
- El operador 5 tiene menor porcentaje de pérdida trabajando en la tarde en comparación a los turnos: mañana y noche.

## CAPÍTULO III

### 3. PROPUESTA DE DISEÑO DE MEJORA

La introducción de este capítulo es el caso práctico de aplicación de las interacciones mencionadas en el capítulo anterior, las que están representadas por los valores de pérdidas obtenidas en el análisis de la situación actual del proceso.

En la siguiente tabla podemos determinar la pérdida promedio obtenida por los operadores en cada uno de sus turnos:

#	TURNO	OPERADOR	PÉRDIDAS (%)
1	MAÑANA	OPERADOR 1	5,62%
3	TARDE	OPERADOR 1	5,46%
2	NOCHE	OPERADOR 1	4,95%
4	MAÑANA	OPERADOR 2	6,84%
6	TARDE	OPERADOR 2	5,65%
5	NOCHE	OPERADOR 2	6,68%
7	MAÑANA	OPERADOR 3	7,73%
9	TARDE	OPERADOR 3	4,52%
8	NOCHE	OPERADOR 3	2,42%
10	MAÑANA	OPERADOR 4	6,31%
12	TARDE	OPERADOR 4	7,06%
11	NOCHE	OPERADOR 4	8,76%
13	MAÑANA	OPERADOR 5	6,41%
15	TARDE	OPERADOR 5	4,22%
14	NOCHE	OPERADOR 5	6,70%

*Tabla III-1 Pérdidas promedio de operadores entre turnos (Paneladora 4)*

*Fuente: Estudio efectuado por los autores*

### **3.1 PLANTEAMIENTO DE TURNOS DE TRABAJO**

La implementación de la mejora se basará específicamente en la proyección de los turnos de trabajo considerando realizar programaciones a los turnos en la que cada operador obtenga los menores promedios de pérdidas.

Para el siguiente cronograma de turno identificaremos a cada operador de la siguiente manera:

<b>Nombre</b>	<b>Abreviatura</b>
OPERADOR 1	<b>ED</b>
OPERADOR 2	<b>JT</b>
OPERADOR 3	<b>JG</b>
OPERADOR 4	<b>RA</b>
OPERADOR 5	<b>SH</b>

*Tabla III-2 Cuadro de operadores (Paneladora 4)*

*Fuente: Estudio efectuado por los autores*

### **3.2 PROGRAMACIÓN SEMANAL DE TURNOS DE TRABAJO**

De acuerdo a los datos obtenidos en la tabla N° 1, se procederá a realizar la selección de los operadores de la línea de producción para la cobertura de los tres turnos establecidos por la organización.



Semana 1							
Turno	Lun.	mar	Mié.	Jue.	Vie.	Sáb.	Dom.
Mañana	ED	ED	ED	ED	ED	ED	ED
Tarde	SH	SH	SH	SH	SH	SH	SH
Noche	JG	JG	JG	JG	JG	N/A	N/A
Mañana	5,62%	5,62%	5,62%	5,62%	5,62%	5,62%	5,62%
Tarde	4,22%	4,22%	4,22%	4,22%	4,22%	4,22%	4,22%
Noche	2,42%	2,42%	2,42%	2,42%	2,42%	N/A	N/A

Promedio de pérdida	4,26%
---------------------	-------

**Tabla III-3 Propuesta semana 1 (Paneladora 4)**

*Fuente: Estudio efectuado por los autores*

Semana 2							
Turno	Lun.	Mar.	Mié.	Jue.	Vie.	Sáb.	Dom.
Mañana	SH	SH	SH	SH	SH	SH	SH
Tarde	JG	JG	JG	JG	JG	JG	JG
Noche	ED	ED	ED	ED	ED	N/A	N/A
Mañana	6,41%	6,41%	6,41%	6,41%	6,41%	6,41%	6,41%
Tarde	4,52%	4,52%	4,52%	4,52%	4,52%	4,52%	4,52%
Noche	4,95%	4,95%	4,95%	4,95%	4,95%	N/A	N/A

Promedio de pérdida	5,33%
---------------------	-------

**Tabla III-4 Propuesta semana 2 (Paneladora 4)**

*Fuente: Estudio efectuado por los autores*

Semana 3							
Turno	Lun.	mar	Mié.	Jue.	Vie.	Sáb.	Dom.
Mañana	ED	ED	ED	ED	ED	ED	ED
Tarde	SH	SH	SH	SH	SH	SH	SH
Noche	JG	JG	JG	JG	JG	N/A	N/A
Mañana	5,62%	5,62%	5,62%	5,62%	5,62%	5,62%	5,62%
Tarde	4,22%	4,22%	4,22%	4,22%	4,22%	4,22%	4,22%
Noche	2,42%	2,42%	2,42%	2,42%	2,42%	N/A	N/A

Promedio de pérdida	4,26%
---------------------	-------

**Tabla III-5 Propuesta semana 3 (Paneladora 4)**  
*Fuente: Estudio efectuado por los autores*

Semana 4							
Turno	Lun.	Mar.	Mié.	Jue.	Vie.	Sáb.	Dom.
Mañana	SH	SH	SH	SH	SH	SH	SH
Tarde	JG	JG	JG	JG	JG	JG	JG
Noche	ED	ED	ED	ED	ED	N/A	N/A
Mañana	6,41%	6,41%	6,41%	6,41%	6,41%	6,41%	6,41%
Tarde	4,52%	4,52%	4,52%	4,52%	4,52%	4,52%	4,52%
Noche	4,95%	4,95%	4,95%	4,95%	4,95%	N/A	N/A

Promedio de pérdida	5,33%
---------------------	-------

**Tabla III-6 Propuesta semana 4 (Paneladora 4)**  
*Fuente: Estudio efectuado por los autores*

<b>Período</b>	<b>Promedio</b>
Semana 1	4,26%
Semana 2	5,33%
Semana 3	4,26%
Semana 4	5,33%
<b>Promedio General</b>	<b>4,80%</b>

*Tabla III-7 Promedio de pérdida mensual estimado*

*Fuente: Estudio efectuado por los autores*

### **Reducción estimada**

Los límites de porcentajes de pérdidas promedio obtenida en la proyección de turnos de trabajo en comparación al promedio inicial de la situación actual es la siguiente:

**Reducción** = Valor Promedio Inicial – Valor Promedio Proyectado

**Reducción** = 6,24 - 4,80

**Reducción** = 1,44

**Porcentaje Reducción** = 23,07 %

## **CONCLUSIONES**

- Mediante la aplicación del diagrama de cajas en la información de pérdidas, se pudo determinar con mayor facilidad que el 17 % de 984 datos se muestran como atípicos generados en cada una de las mediciones realizadas, los mismos que son puntuales y no se repiten en el proceso.
- Se demuestra también que el 50 % de los datos analizados tienen un promedio de pérdidas disperso entre el rango de 2,9 % y 8,7 %.
- Se prueba mediante el análisis estadístico descriptivo cuál es el comportamiento real del proceso considerando la relación con los operadores y turnos de trabajo.
- Al realizar el análisis de la influencia de la variable turno en las pérdidas, encontramos que en el turno de la mañana se produce mayor pérdida, en la noche menos que la mañana, pero en la tarde es el mejor turno debido a que se produce menos pérdida durante la producción.
- Se observa también que al realizar el análisis de la influencia de la variable operador existen diferencias en el promedio de pérdida entre cada uno de ellos.
- Podemos observar que en promedio el operador 1 produce una menor pérdida en sus turnos de trabajo a diferencia del operador 4 que produce un mayor promedio de pérdida de todo el grupo analizado.
- Por otro lado, se opta por la aplicación del análisis de variables con interacciones, ayudando a determinar cuál es el mejor rendimiento de cada uno de los operadores en los diferentes turnos. Siendo el objeto de la medición, el generar la menor cantidad de pérdidas.
- Se observa que el operador 1 tiene menos pérdida en los tres turnos; el operador 2 tiene más pérdida que el operador 1 pero es constante en los tres turnos; el operador 3 tiene menos pérdida en la noche que en la tarde y mañana; el operador 4 tiene menos pérdida en la mañana que en la tarde y noche; el operador 5 tiene menos pérdida en la tarde que en la mañana y noche.
- Se demuestra mediante el planteamiento teórico de esquemas de turnos de trabajo, cuál sería el promedio de pérdida mensual estimada en la fabricación de cubiertas onduladas. Demostrando una reducción de 23,07 puntos porcentuales del valor inicial de pérdida promedio.

## **RECOMENDACIONES**

- Establecer un mecanismo para el análisis de competencias e implementar test psicométricos a operadores de líneas, a fin de determinar cuál es el factor influyente que marca diferencias en sus resultados.
- Realizar revisión y actualización de métodos de operación de equipos, con la finalidad de establecer un estándar de operación.
- Realizar análisis de componentes químicos y mecánicos de la materia prima que se utiliza para la fabricación de cubiertas onduladas previo a la compra, para de esta manera descartar su incidencia en la generación de pérdidas en el proceso.
- Implementar plan de mantenimiento predictivo y de mejoras en la línea de producción, para de esta manera facilitar la operación en la fabricación de la cubierta.
- Realizar nuevo cálculo de la meta o límite máximo de pérdidas mediante un análisis estadístico, tomando los valores reales obtenidos en los últimos años.

## BIBLIOGRAFÍA

- [1] **Barrio Vilar, José.** *Control estadístico de los procesos.* España : Fundación Confemetal, 2005.
- [2] **Pedro, Sanchez Algarra.** *"Métodos Estadísticos Aplicados"*. España : s.n., 2006.
- [3] **Mario, Suarez Ibujez.** "Medidas de Dispersión y simetría de Datos". [En línea] Internet. [Citado el: 29 de Marzo de 2012.] [www.monografias.com/trabajos89/medidas-de-dispersion/medidas-de-dispersion.shtml](http://www.monografias.com/trabajos89/medidas-de-dispersion/medidas-de-dispersion.shtml) .
- [4] **Juran J.M./Gryna Franc Jr./Bingham R.S.** "Manual de Control de Calidad" Volumen 2 Segunda edición, Editorial Reverté 2005.
- [5] **Lind Douglas A., Marchal William G., Wathen. Samuel A.** *"Estadística Aplicada a los Negocios y la Economía"*. Decimotercera Edición.
- [6] **Ley del Sistema Ecuatoriano de la Calidad, Legislación Conexa** "Corporación de Estudio y Publicaciones-CEP "Año de publicación 2007.
- [7] **Alcalá Gustavo** "Análisis Simple de la Varianza"[En línea]Internet [www.monografias.com/trabajos61/analisis-varianza/analisis-varianza2.shtml](http://www.monografias.com/trabajos61/analisis-varianza/analisis-varianza2.shtml)
- [8] Comité del ISO /CASCO Norma ISO/ IEC 9001:2008
- [9] **Instituto de Normalización Ecuatoriano** Reglamento Técnico Ecuatoriano INEN 2221:2008
- [10] **J.C., Miller N.J y Miller.** *Estadística y Quimiometría para Química Analítica.* s/n : Cuarta Edición, s/n.
- [11] **Software, Desarrolladores de.** Software Estadístico Minitab. [En línea] [Citado el: 30 de Marzo de 2012.] [www.minitab.com](http://www.minitab.com).
- [12] **Desarrolladores de Software.** [En línea] RStudio Ink. [Citado el: 30 de Marzo de 2012.] [www.rstudio.org](http://www.rstudio.org).