

# **ESTADISTICA**

**UNA APROXIMACIÓN AL  
ENFOQUE POR COMPETENCIAS**

## **CAPITULO 1**

### **INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA**

**Curso de Estadística 2016-1  
Pontificia Universidad Javeriana Cali  
Facultad de Ingeniería  
Departamento de Ciencias Naturales y Matemáticas**

**Autores: Sandra Milena Ramírez, María del Pilar Marín,  
Guillermo Valdés, Gerardo Ramírez y Daniel Enrique González**

# Índice general

<b>1. Estadística descriptiva</b>	<b>3</b>
1.1. Introducción . . . . .	5
1.2. Conceptualización . . . . .	5
1.2.1. Planteamiento del problema. . . . .	6
1.2.2. El marco teórico. . . . .	6
1.2.3. Alcance del estudio. . . . .	6
1.2.4. Variables de interés. . . . .	7
1.2.5. Diseño del estudio. . . . .	9
1.2.6. Conceptos básicos en Estadística Descriptiva . . . . .	10
1.2.7. Análisis de datos. . . . .	10
1.2.8. Tipos de estudio. . . . .	10
1.2.9. Conclusiones y recomendaciones. . . . .	11
1.3. Tablas de distribución de frecuencia . . . . .	11
1.3.1. Variable cualitativa . . . . .	12
1.3.2. Variable cuantitativa . . . . .	12
1.3.3. Tabla de contingencia . . . . .	13
1.4. Resumen numérico . . . . .	13
1.4.1. Indicadores de posición. . . . .	16
1.4.2. Indicadores de posición central . . . . .	16
1.4.3. Indicadores de dispersión . . . . .	17
1.4.4. Indicadores de forma . . . . .	18
1.5. Resumen gráfico . . . . .	20
1.5.1. Gráfico de torta . . . . .	20
1.5.2. Diagrama de barras . . . . .	20
1.5.3. Histograma . . . . .	23
1.5.4. Polígono de frecuencia . . . . .	24
1.5.5. Ojiva . . . . .	24
1.5.6. Diagrama de tallo y hojas . . . . .	25
1.5.7. Diagrama de dispersión . . . . .	26
1.5.8. Diagrama de caja y bigotes . . . . .	26
1.5.9. Diagrama de datos longitudinales o series temporales . . . . .	27
1.6. Códigos en R-Project . . . . .	29
1.6.1. Variables Cualitativas . . . . .	29
1.6.2. Variables Cuantitativas . . . . .	29
1.7. Ejercicios resueltos . . . . .	31
1.7.1. . . . .	31

1.7.2.	. . . . .	34
1.7.3.	. . . . .	36
1.7.4.	. . . . .	37
1.8. Problemas Propuestos	. . . . .	42
1.8.1.	. . . . .	42
1.8.2.	. . . . .	42
1.8.3.	. . . . .	43
1.8.4.	. . . . .	46
1.8.5.	. . . . .	46
1.8.6.	. . . . .	46
1.8.7.	. . . . .	47
1.8.8.	. . . . .	48
1.8.9.	. . . . .	49

# Capítulo 1

## Estadística descriptiva

### Competencias:

El presente capítulo busca contribuir al desarrollo de las competencias presentadas a continuación atendiendo los indicadores correspondientes.

#### 1. Comunicar efectivamente.

- C1. El alumno identifica y usa la nomenclatura estadística adecuada presentada en el capítulo para representar ideas y formular preguntas asociadas con la información presente en un problema objeto de estudio.
- C2. El alumno selecciona el procedimiento estadístico a utilizar de acuerdo al tipo de datos asociados con un problema de estudio.
- C3. El alumno interpreta los resultados encontrados en el análisis de los datos en el contexto del problema involucrado.
- C4. El alumno interpreta los resultados encontrados en el análisis de los datos en el contexto del problema involucrado.

#### 2. Resolver problemas en contexto.

- R1. El alumno identifica el tipo de software que debe utilizar para resolver el problema bajo estudio y las respuestas que espera obtener.
- R2. El alumno estará en capacidad de emplear el software estadístico para evaluar las respuestas que espera obtener.
- R3. El alumno identifica cuáles son los indicadores estadísticos adecuados para conocer las características especiales del conjunto de datos, tales como: posición, dispersión y forma, que le permiten resolver el problema objeto de estudio.
- R4. El estudiante distingue adecuadamente las características presentes en un conjunto de datos mediante tablas de frecuencia, gráficas e indicadores de tendencia central, dispersión, forma y posición.
- R5. El alumno selecciona el tipo de indicadores que debe utilizar para caracterizar la información relacionada con la información involucrada en los problemas.
- R6. El alumno, mediante un proceso de simulación valida las respuestas obtenidas para la solución de un problema.

3. Manejar herramientas computacionales.

Cp1. El alumno utiliza algunas herramientas computacionales que le permiten dar solución al problema de acuerdo al tipo de datos involucrados en el mismo

**Referencia:**

- (1) Walpole, Myers, Myers y Ye (2007). Probabilidad y Estadística para ingeniería y Ciencias: Pearson ? Prentice Hall. Octava edición (2007).
- (2) Navidi, W., (2006). Estadística para ingenieros y científicos. McGraw Hill.
- (3) Malhotra N. (2008). Investigación de mercados: Pearson.
- (4) Behar Roberto, Yepes Mario (1996). Estadística un enfoque descriptivo. Universidad del Valle.
- (5) Miller y Freund. Probabilidad y estadística para ingenieros. Pearson. Octava edición (2012)

## 1.1. Introducción

En la vida real y particularmente en situaciones relacionadas con ingeniería, se necesita conocer el comportamiento de conjuntos de datos, ya sea existentes o conseguirlos para resolver problemas que se presentan a nivel empresarial.

El presente capítulo busca colaborar a dar solución al problema, resumiendo de forma adecuada la información que se necesita para su estudio.

En este capítulo se presentan aspectos relevantes sobre la metodología estadística. Se brindan los conceptos, su uso y algunas técnicas de aplicación. Estas técnicas serán la base para la presentación de los conceptos teóricos de los capítulos posteriores. Se presentan, además aspectos relevantes de la estadística descriptiva con ejemplos resueltos a partir de la información tomada de una empresa del sector de producción. Luego, se presenta un conjunto de ejercicios resueltos. Al final, se proponen ejercicios tomados de autores referidos en los que los estudiantes podrán ahondar sobre los temas de este capítulo.

## 1.2. Conceptualización

En este capítulo también se aborda el estudio de la estadística descriptiva. Uno de los objetivos de la estadística descriptiva es el de elaborar juicios sobre un conjunto de datos en términos de su localización, dispersión y la forma de la distribución cuando se agrupan los datos. Inicialmente se presentan conceptos propios de una investigación, donde cobra sentido el uso de los métodos estadísticos, luego se presentan las tablas, se analizan los indicadores de posición, dispersión y forma, y las gráficas más utilizadas.

Es importante iniciar el estudio de la estadística descriptiva estableciendo un conjunto de pasos secuenciales que involucran generalmente:

1. Planteamiento del problema.

2. Definición de un marco teórico.
3. Visualización del alcance del estudio.
4. Definición de hipótesis y variables.
5. Diseño del estudio.
6. Definición, selección de la muestra y recolección de la información.
7. Procesamiento y análisis de datos.
8. Tipos de estudio y reporte de resultados.
9. Conclusiones y recomendaciones.

Aunque en el presente capítulo solo se consideran algunos de estos aspectos mencionados

### **1.2.1. Planteamiento del problema.**

El problema se define a partir del análisis o de la observación crítica de una situación. En este sentido, se cuenta con algunas técnicas estadísticas para detectar anomalías o situaciones críticas que llevan a identificar y definir el problema. En la distribución de los chocolates producidos en una fábrica por ejemplo, pudiera observarse una disminución en las ventas. Esta situación conduce hacia un seguimiento dentro del proceso para detectar la raíz del problema.

### **1.2.2. El marco teórico.**

Se construye a partir de la revisión exhaustiva de la literatura acerca del problema y a partir de esto, se definen los objetivos, alcances de la investigación y se elaboran las hipótesis estadísticas que guiarán la intención transformadora de una realidad hipotética. Las hipótesis constituyen afirmaciones o conjeturas sobre una o varias características de la población, si estas características son numéricas, son llamadas parámetros. Por ejemplo, en el estudio del control de las especificaciones en la producción de chocolate en barras, algunas hipótesis de investigación podrían ser:

- El peso promedio de las barras de chocolate producidas no está cumpliendo con estándares propuestos.
- El sabor no satisface a un porcentaje significativo de clientes.

### **1.2.3. Alcance del estudio.**

Un estudio estadístico puede hacerse siguiendo dos enfoques: Descriptivo, si el objetivo busca la caracterización de una o más variables involucradas. El estudio tendrá alcance inferencial si el objetivo tiende a describir un conjunto mayor (población) a partir de un subconjunto (muestra) de elementos. Para este fin será necesario apoyarse en la teoría de muestreo.

### 1.2.4. Variables de interés.

Un examen sobre la información requerida para el cumplimiento de los objetivos lleva a identificar características medibles que deben cumplir especificaciones y por tanto dan origen a hipótesis sobre los parámetros que las representan. Por ejemplo, en los enunciados anteriores están involucradas respectivamente las variables:

- Número de pedidos/semana devueltos.
- Peso de las barras de chocolate.
- Preferencias de sabor de los consumidores

Dado que una variable puede tener distinta naturaleza en la explicación de los fenómenos, es preciso clasificarlas para facilitar el uso de los métodos y dar rigor al estudio.

#### 1.2.4.1. Tipo de variables

Las variables pueden clasificarse según su origen, su naturaleza o su relación con otras. La Figura 1.1 ilustra la clasificación de variables de acuerdo a lo expresado anteriormente.



Figura 1.1: Clasificación de variables.

**1.2.4.1.1. Por su origen:** Una variable es aleatoria cuando los valores resultantes de una medición no se pueden predecir de antemano. Podría decirse que ese valor se desconoce por completo. Si antes de medir, puede predecirse el valor que tendrá la variable, entonces se dice que ésta tiene carácter determinístico. Por ejemplo, en un proceso de producción de un artículo en grandes lotes, no se puede predecir el número de artículos defectuosos que se obtendrán, por esta razón se dice que el número de artículos defectuosos es una variable aleatoria. Las variables determinísticas aparecen fundamentalmente en relaciones comunes en física y química. Aunque vale la pena aclarar que estas variables son aproximadamente determinísticas.



**1.2.4.1.2. Por su naturaleza:** Una variable es cuantitativa si denota características cuyos valores pertenecen al conjunto de los reales, tales como edad, peso y el sueldo de los empleados o el tiempo de vida de un equipo electrónico. Estas variables cuantitativas, a la vez se pueden clasificar en discretas o continuas.

La variable es discreta cuando el conjunto de valores que toma la variable es finito o infinito numerable. Por ejemplo, el número de hijos por familia, el número de fallas que presenta un aparato electrónico en una semana.

La variable es continua cuando el conjunto de valores que puede tomar la variable corresponde a un intervalo o a la unión de intervalos. Por ejemplo, el peso de una persona, esta variable puede tomar cualquier valor en un intervalo de pesos determinado, por ejemplo, entre 40 y 90 kilos.

Una variable es cualitativa si los valores representan categorías tales como estado civil, profesión y nacionalidad de los empresarios, marca, estado y tipo de defecto de un equipo.

**1.2.4.1.3. Por su relación con otras variables:** Es frecuente que un estudio tenga por objetivo describir el comportamiento de una variable específica a partir de otras variables. Por ejemplo, el tiempo de vida de un auto puede informar sobre el número de visitas por año al taller. En este caso, se habla de variables dependientes e independientes o explicativas. En el Tabla 1.1 se detalla la clasificación de las variables para el caso de producción de chocolate.

Tabla 1.1: Variables producción hipotética de chocolate

Variable	Origen	Naturaleza	Relación con otras
Número de pedidos devueltos / semana	Aleatorio	Cuantitativa discreta	Depende del peso y preferencia
Peso de las barras de chocolate	Aleatorio	Cuantitativa continua	Explica el número de pedidos devueltos
Preferencias calóricas de los consumidores	Aleatorio	Cualitativa	Explica el número de pedidos devueltos

#### 1.2.4.2. Escalas de medición

Para efectos de presentación de los resultados de un estudio, es importante establecer los niveles o escalas de medición de las variables. Se cuenta con 4 niveles, como se muestra a continuación

**1.2.4.2.1. Escala nominal:** Los valores son categorías no comparables, sin presencia de orden distinta al alfabético entre ellos. Las relaciones posibles entre los valores son ser iguales o ser diferentes. Ejemplo: género, estado civil.

**1.2.4.2.2. Escala ordinal:** Los valores son comparables, en el sentido de que se pueden ordenar y la variable es cualitativa. Ejemplo: Calidad de servicio de facturación medida con las alternativas excelente, bueno, regular, malo y pésimo.

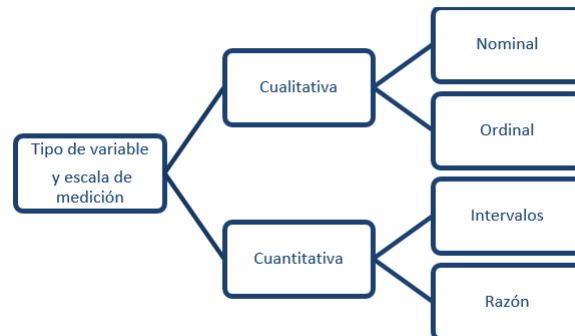


Figura 1.2: Escalas de medición.

**1.2.4.2.3. Escala de intervalo:** Se pueden medir variables cuantitativas las cuales no solo tienen en cuenta un orden entre sus valores como en la escala ordinal, sino que involucra el concepto de unidad de distancia, la cual permite establecer la diferencia entre dos medidas y esta diferencia tiene sentido en la medida que se está considerando. Requiere de un punto de referencia o cero relativo, el cual es arbitrario y no indica ausencia de lo que se está midiendo. Por ejemplo, las escala de grados centígrados y grados Fahrenheit, con las que se acostumbra medir la temperatura tienen diferentes temperaturas cero y diferentes definiciones de grado o unidad de medida.

**1.2.4.2.4. Escala de razón:** Pueden medirse variables cuantitativas que cuentan con un cero absoluto (ausencia de lo que está midiendo) y cuyos valores, además de poder establecer la diferencia entre dos medidas, se pueden relacionar a través de la razón entre sus valores: Ejemplo: el peso de una persona, utilidad de una empresa, nivel de endeudamiento de una empresa. En el ejemplo de producción de chocolate en barras, el número de pedidos devueltos /semana y el peso de las barras estarían en escala de razón y las preferencias en el nivel nominal.

## 1.2.5. Diseño del estudio.

Esta fase incluye la especificación de factores que complementan el estudio para el logro de objetivos como el tipo de estudio, estrategias de muestreo, presupuesto y cronogramas entre otros. A nivel de ingeniería es muy frecuente el estudio de tipo experimental, donde el investigador determina intencionalmente algunos valores de las variables y tiene el control sobre otras. En este estudio se busca la incidencia que tienen los diferentes niveles de las variables sobre la variable dependiente. Otro estudio es el no experimental, donde no se controlan las variables que se investigan solamente se recogen en su “estado natural”.

## 1.2.6. Conceptos básicos en Estadística Descriptiva

### 1.2.6.1. Población.

A nivel descriptivo se utilizará inicialmente como definición de población al conjunto de todos los elementos sobre los cuales recaen las conclusiones de un estudio estadístico. Ejemplo: En el estudio sobre el rendimiento estudiantil en una institución, la población está conformado por todos estudiantes matriculados en la institución. En el análisis del porcentaje de defectos de un proceso, toda la producción del proceso representa la población.

Posteriormente en el capítulo seis de este texto se da un enfoque más detallado del concepto de población, para hacerla de aplicación más estadística

### 1.2.6.2. Muestra.

Subconjunto de elementos de una población.

## 1.2.7. Análisis de datos.

### 1.2.7.1. Datos

Se considera que un dato es la observación de la medida de una variable tomada con un instrumento definido para establecer su valor.

Los datos pueden corresponder a estudios:

- **Transversales:** Cuando se recogen en un periodo corto de tiempo y reflejan lo que ocurre en un momento dado.
- **Longitudinales:** Cuando se refiere a la observación de un grupo a lo largo del tiempo.

### 1.2.7.2. Procesamiento de los datos

Esta etapa corresponde al conjunto de procedimientos que facilitan el análisis de la información para sustentar las interpretaciones y conclusiones en un estudio. Se incluyen tablas, gráficos, indicadores numéricos de una y varias variables, para lo cual es importante disponer de una base de datos que permita el uso de herramientas computacionales o paquetes estadísticos. Entre los más usados, sin jerarquía de uso están: SPSS, Stata, Minitab, SAS, Statgraphics , R, o también paquete no especializado como EXCEL

## 1.2.8. Tipos de estudio.

### 1.2.8.1. Estadística descriptiva.

Generación de tablas de distribución, gráficos, indicadores de centro, dispersión, forma y posición que faciliten la comprensión e interpretación de los datos.

Una vez definido un estudio y recolectada la información, el paso a seguir es el procesamiento de los datos que se inicia con el diseño de una base de datos con un registro por cada individuo

o elemento encuestado u observado. Luego de esto, un paso obligado es la depuración de datos, con esto se busca detectar errores en la lectura o transcripción de datos, para lo cual los gráficos son de excelente ayuda. El procesamiento se inicia resumiendo los datos lo cual se puede obtener con tablas, gráficos o indicadores cuantitativos

#### **1.2.8.2. Inferencia estadística.**

Estudia los métodos para sacar conclusiones sobre propiedades de una población a partir de propiedades una muestra aleatoria. Se realiza a través de la estimación (puntual o por intervalos de confianza) de parámetros o mediante prueba de hipótesis (paramétrica o no paramétrica). Esta etapa será considerada más adelante.

#### **1.2.9. Conclusiones y recomendaciones.**

La parte final está relacionada con la elaboración de reportes de resultados para la toma de decisiones o para ser divulgados.

Como se mencionó inicialmente, la parte central de este capítulo corresponde a la estadística descriptiva, para lo cual se analizarán los valores de algunas variables correspondientes a una empresa de producción de material de protección sanitaria femenino. Las variables que se considerarán son:

- Pesos de los cojines protectores (en gramos).
- Volumen de aceite(ml)
- Longitud de las cintas adhesivas (cms).
- Número de empaques con sellado defectuoso por muestra de 800.
- Número de quejas por lote de producción.
- Tipos de defectos de los cojines.

Con el ánimo de evaluar el proceso, se realizará un análisis descriptivo de algunas de estas variables, teniendo en cuenta el tipo de variable y los aspectos teóricos involucrados en cada etapa del proceso. Este análisis comprende: Tablas de frecuencia, gráficos e indicadores numéricos.

### **1.3. Tablas de distribución de frecuencia**

Las tablas de distribución de frecuencias se utilizan tanto para datos cualitativos como cuantitativos y permiten ver de una manera resumida la información que hay en los datos.

### 1.3.1. Variable cualitativa

Si la variable es cualitativa, la tabla muestra las categorías de la variable en una columna y las frecuencias absolutas (conteos) y relativas (fracciones o porcentajes), en otras columnas. Para el caso de la variable “tipo de defecto en artículos producidos”, la cual es cualitativa y tiene seis categorías, la distribución de frecuencias se presenta en la Tabla 1.2

Tabla 1.2: Distribución de frecuencia para tipo de defecto.

<b>Tipo de defecto</b>	<b>Frecuencia</b>	<b>Frecuencia relativa</b>
Crítico	3	0.03
Crítico y menor	1	0.01
Mayor	7	0.07
Mayor y menor	2	0.02
Menor	11	0.11
Ninguno	76	0.76

En la Tabla 1.2 se puede observar que de un total de 100 artículos examinados, el 76 % de ellos no muestra ningún tipo de defecto y el defecto que se presenta con mayor frecuencia es el tipo de defecto menor (11 %), seguido del tipo de defecto mayor (7 %).

### 1.3.2. Variable cuantitativa

Si la variable es cuantitativa, la tabla muestra en dos columnas los límites de las clases o grupos en que se han clasificado los datos numéricamente ordenados, el punto medio de cada clase, las frecuencias absolutas y las frecuencias relativas, en forma similar que las cualitativas, pero además muestran las frecuencias absolutas acumuladas y las frecuencias relativas acumuladas. Con frecuencia las clases son del mismo ancho (diferencia entre el límite superior y el límite inferior). Cuando las clases no son del mismo ancho, es necesario generar otra columna correspondiente a la densidad de frecuencia relativa, la cual se obtiene dividiendo la frecuencia relativa de la clase entre el ancho de la clase. Esta densidad ayuda a identificar como es la concentración de los datos y es fundamental para la representación gráfica de los mismos.

La tabla de frecuencia para la variable peso de los cojines, que es una variable cuantitativa, medida en gramos y correspondiente a la escala de razón, se presenta en la Tabla 1.3.

Tabla 1.3: Tabla de frecuencia para la variable peso de los cojines.

	<i>Límite</i>	<i>Límite</i>			<i>Frecuencia</i>	<i>Frecuencia</i>	<i>Frecuencia</i>
<i>Clase</i>	<i>Inferior</i>	<i>Superior</i>	<i>Punto Medio</i>	<i>Frecuencia</i>	<i>Relativa</i>	<i>Acumulada</i>	<i>Rel. Acum.</i>
1	4.7	4.8	4.75	10	0.1	10	0.1
2	4.8	4.9	4.85	13	0.13	23	0.23
3	4.9	5	4.95	30	0.3	53	0.53
4	5	5.1	5.05	26	0.26	79	0.79
5	5.1	5.2	5.15	16	0.16	95	0.95
6	5.2	5.3	5.25	3	0.03	98	0.98
5	5.3	5.4	5.35	2	0.02	100	1

Tabla 1.4: Distribución de microprocesadores por proveedor y velocidad

	<b>Proveedor1</b>	<b>Proveedor2</b>	<b>Proveedor3</b>	<b>Proveedor4</b>
Baja	26	23	15	32
Media	118	93	116	121
Alta	56	84	69	47

En esta tabla se puede observar cómo están distribuidos los pesos de los cojines. Para la empresa productora es muy importante conocer esta distribución puesto que los pesos deben estar concentrados alrededor de 5 gramos y mediante este resumen se pretende descubrir si hay algún inconveniente con el llenado de estos cojines. Vale resaltar que entre los datos analizados, hay 10 % de cojines con pesos entre 4.7 y 4.8 gramos y un 5 % de cojines con pesos entre 5.2 y 5.4 gramos.

### 1.3.3. Tabla de contingencia

Para resumir de manera simultánea los datos de dos variables, o para analizar la relación de dependencia o independencia entre dos variables, generalmente cualitativas, es necesario estudiar su distribución conjunta o tabla de contingencia.

La tabla de contingencia es una tabla de doble entrada, filas y columnas, donde en cada casilla o celda figura el número de casos o individuos que poseen un nivel específico de uno de los factores o características analizadas y un nivel específico del otro factor analizado.

En este tipo de tabla, se puede mostrar la frecuencia relativa de los datos de una celda con respecto al total de la muestra, al total de la fila ó al total de la columna.

#### Ejemplo

Un grupo de ingenieros, interesados en verificar la calidad de los microprocesadores suministrados por cuatro proveedores, selecciona una muestra de 200 microprocesadores de cada uno de los proveedores y los clasifica de acuerdo a su velocidad de procesamiento. en la Tabla 1.4 se muestra los datos obtenidos fruto de la clasificación realizada.

A partir de la Tabla 1.4 anterior se puede afirmar que el 12 % de los microprocesadores analizados son de baja velocidad, el 56 % de velocidad media y el 32 % son de alta velocidad. Se puede afirmar también que el proveedor que suministró mayor porcentaje de microprocesadores de alta velocidad fue el proveedor 2 (33 %) y el que suministró menor porcentaje de microprocesadores de alta velocidad fue el proveedor 4 (18 %), ver Tabla 1.5

## 1.4. Resumen numérico

En estudios cuantitativos además de las tablas y del resumen gráfico que se hará posteriormente, se precisa de análisis más finos que unidos a las concentraciones de datos, soporten las estrategias o decisiones en la interpretación o solución del problema. Las medidas

Tabla 1.5: Distribución porcentual de microprocesadores de cada proveedor respecto al total en cada una de las calidades

	<b>Proveedor1</b>	<b>Proveedor2</b>	<b>Proveedor3</b>	<b>Proveedor4</b>	<b>Total</b>
Baja	27	24	16	33	100
Media	26	21	26	27	100
Alta	22	33	27	18	100

permiten cuantificar algunas características en los datos. Son frecuentes las que generan un valor que representa al conjunto, también llamadas de posición central; las que miden la variabilidad de los datos, llamadas de dispersión y otras que describen la posición que ocupan algunos valores en la distribución, conocidas como estadísticos de orden, dentro de los cuales se tienen los percentiles. Algunos modelos o fórmulas para obtener los distintos estadísticos se presentan resumidos en el Tabla 1.6.

Tabla 1.6: Fórmulas para indicadores más comunes

Modelo	Indicador
$\bar{x} = \frac{\sum x_i}{n}$	Promedio de datos en la lista
$\bar{x} = \frac{\sum x'_i n_i}{n}$	Promedio para datos Agrupados
$\bar{x}_w = \frac{\sum x_i w_i}{\sum w_i}$	Promedio ponderado
$Me = \begin{cases} x_{(n+1)/2}, & n \text{ impar} \\ \frac{x_{n/2} + x_{(n+2)/2}}{2}, & n \text{ par} \end{cases}$	Mediana de datos en lista
$P_k = L_{i-1} + \left( \frac{k}{100} - F(L_{i-1}) \right) C_i / f_i$	Percentil $K$ . En caso de mediana $K=50$
$F(x) = F(L_{i-1}) + \frac{f_i}{C_i}(x - L_{i-1})$	Frecuencia relativa acumulada hasta un punto $x$
$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$	Varianza muestral para datos en lista
$s^2 = \frac{\sum_{i=1}^m n_i (x'_i - \bar{x})^2}{n-1}$	Varianza muestral para datos agrupados
$s = \sqrt{s^2}$	Desviación estándar
$CV = \frac{s}{\bar{x}} 100 \%$	Coefficiente de variación
$f_i^* = \frac{f_i}{C_i}$	Densidad de frecuencia relativa
$p = \frac{\text{Casos favorables}}{\text{Total de casos}}$	Proporción de éxitos

De la Tabla 1.6,

$K$  : Percentil buscado. En la mediana,  $K = 50$ ,  $P_{50}$ .

$C_i$  : Ancho del intervalo  $i$ .

$x'_i$  : Marca de clase o punto medio de la clase  $i$ .

$n_i$  : Frecuencia absoluta de la clase  $i$ .

$w_i$  : Factor de ponderación.

$m$  : Número de intervalos o clases.

$f_i$  : Frecuencia relativa del intervalo  $i$ .

$F_i$  : Frecuencia relativa acumulada hasta el intervalo  $i$ .

$L_{i-1}$  : Limite inferior del intervalo que contiene el indicador de interés.

$f_i^*$  : Densidad de frecuencia relativa.



### 1.4.1. Indicadores de posición.

Estos indicadores dan información sobre la posición de un valor con respecto al grupo o muestra. Los más comunes son la mediana, valor que divide la distribución en dos partes, cada una de 50 % ( $Me$ ). Los cuartiles, tres valores que dividen al grupo de datos en cuatro partes de igual porcentaje (25 % cada una):  $Q_1$ ,  $Q_2$  y  $Q_3$ . Los deciles, nueve números que dividen al grupo de datos en diez partes de igual porcentaje, cada una de 10 %:  $D_1$ ,  $D_2$ ,  $D_3$ ,  $D_4$ ,  $D_5$ ,  $D_6$ ,  $D_7$ ,  $D_8$  y  $D_9$ . Los percentiles, noventa y nueve números que dividen al grupo de datos en 100 partes cada una de un 1 %:  $P_1$ ,  $P_2$ ,  $P_3$ , ...,  $P_{99}$ .

Se puede observar que:

$$Q_1 = P_{25},$$

$$Me = Q_2 = D_5 = P_{50},$$

$$Q_3 = P_{75}$$

La Tabla 1.7 muestra algunos percentiles para la variable volumen de aceite

Tabla 1.7: Percentiles para la variable volumen.

$k$	Percentil $P_k$
1 %	195.270
5 %	195.565
10 %	195.905
25 %	197.500
50 %	200.050
75 %	202.275
90 %	204.000
95 %	204.7
99 %	204.97

De la tabla anterior se podría decir lo siguiente:

- El 10 % de los datos de volumen de aceite son menores o iguales a 195.905 mililitros
- El 50 % de los datos de volumen de aceite son menores o iguales a 200.05 mililitros.
- El 75 % de los datos de volumen de aceite son menores o iguales a 202.275 mililitros.
- El 95 % de los datos de volumen de aceite son menores o iguales a 204.7 mililitros

### 1.4.2. Indicadores de posición central

Como indicadores de posición central se pueden citar: la media aritmética, la mediana, el rango medio, la media truncada, la media geométrica, la media armónica y en algunos casos la moda o dato más frecuente. Cada uno de estos brinda un valor representante de todo el grupo de datos.

Aunque generalmente se considera la media aritmética como un buen indicador de posición central, cuando los datos tienen un comportamiento muy asimétrico, la media no es un buen representante de los datos, en estos casos es mejor utilizar la mediana, la cual sería mejor representante de los datos ya que no se ve afectada por la presencia de datos extremos (muy grandes o muy pequeños comparados con el resto de los datos).

En la Tabla 1.8, se presentan algunos indicadores de tendencia central para las variables: peso de los cojines, volumen del aceite, longitud de las cintas adhesivas, número de empaques con sellado defectuoso y número de quejas por lote de producción.

Tabla 1.8: Indicadores de posición central para variables tomadas de una empresa productora de elementos de aseo femenino.

	Media ( $\bar{x}$ )	Mediana( $Me$ )	Moda( $Mo$ )
<b>Peso</b>	4.9975	4.992	
<b>Vol</b>	199.87	200.05	
<b>Long</b>	8.264	7.373	
<b>Malsellado</b>	11.94	12	15
<b>Porcentaje de quejas</b>	1.76	2	1

### 1.4.3. Indicadores de dispersión

En el análisis de datos, no basta con conocer un indicador de centro, se requiere de información adicional que permita establecer la confianza con que estos pueden usarse. Por ejemplo, suponga que se tienen información sobre la edad promedio de un grupo de 4 personas (media = 20 años). Con solo esta información no se puede afirmar que el grupo está conformado por jóvenes. El grupo puede estar conformado por personas con edad de: 18, 19, 21 y 22 años. En este caso la media representaría muy bien al grupo. Pero si el grupo corresponde a una familia de padres y sus bebés con edades respectivas de 39, 38, 2 y 1, a pesar de que la media es 20, está lejos de hacer una representación coherente del grupo. Se puede notar que los datos del primer grupo están entre 18 y 22 y los del segundo grupo entre 1 y 39; esta situación permite afirmar que hay mayor dispersión en el segundo grupo de datos. Por este motivo es necesario contar con información adicional sobre la dispersión de los datos.

Los indicadores de dispersión más conocidos son: rango, varianza, desviación estándar o típica, coeficiente de variación (utilizado para comparar varios juegos de datos con diferentes unidades), cuyas fórmulas se encuentran en la Tabla 1.6.

En la Tabla 1.9, se presentan algunos indicadores de dispersión para las variables: peso de los cojines, volumen del aceite, longitud de las cintas adhesivas, número de empaques con sellado defectuoso y número de quejas por lote de producción.

Cuando se quiere comparar la variabilidad de dos grupos de datos, se debe tener en cuenta lo siguiente:

- Si las unidades de medida de las variables que se están midiendo son las mismas y los promedios son iguales, se puede hacer la comparación de las dispersiones mediante la desviación estándar

Tabla 1.9: Indicadores de dispersión para variables tomadas de una empresa productora de elementos de aseo femenino

Variable	Error estándar	Desv.Est.	Varianza	CoefVar
Peso	0.0136	0.1362	0.0186	2.73
Volumen	0.293	2.93	8.57	1.46
Longitud	0.364	3.641	13.259	44.06
Malsellado	0.324	3.241	10.501	27.14
Número de quejas	0.138	1.379	1.901	78.35

- Si las unidades de medida de las variables que se están midiendo no son las mismas, se puede hacer la comparación de las dispersiones mediante el coeficiente de variación, puesto que este carece de unidades.
- Para las variables planteadas en la Tabla 1.9, como las unidades de medida no son las mismas, se recurre al coeficiente de variación, el cual permite concluir que la de menor variabilidad es “Volumen” (1.46 %) y la de mayor variabilidad es “número de quejas” (78.35 %).

De manera acordada generalmente, se tiene la siguiente clasificación:

- Si  $CV \leq 5\%$ , altamente homogéneos.
- Si  $5\% \leq CV \leq 20\%$ , homogéneos.
- Si  $20\% < CV \leq 50\%$ , heterogéneos
- Si  $CV > 50\%$ , altamente heterogéneos

#### 1.4.4. Indicadores de forma

Además de identificar los indicadores de posición y de dispersión, es necesario conocer otros indicadores que sirven para conocer la forma que presentan las distribuciones de los datos.

##### 1.4.4.1. Curtosis

La curtosis se utiliza en distribuciones simétricas (o casi simétricas); indica que tan “puntiaguda” es la distribución de los datos y la comparación se hace con respecto a la forma de una distribución que tiene forma de campana (normal). Se suele medir con el coeficiente de curtosis  $CA$ , donde:

$$CA = \frac{\frac{1}{n} \sum_{i=1}^m (x_i - \bar{x})^4 f_i}{\sigma^4} - 3$$

Si  $CA$  está muy cerca de cero, se establece como básica y recibe el nombre de mesocúrtica o normal. Si  $CA$  es mayor que cero, la distribución es más puntiaguda, hay una mayor concentración de los datos que en la normal y recibe el nombre de leptocúrtica. Si  $CA$  es menor que cero, la distribución es más aplanada, hay una menor concentración de los datos que en la normal y se llama platicúrtica.

### 1.4.4.2. Sesgo o coeficiente de asimetría

Mide el grado de asimetría de los datos respecto a la media. Esta característica, en términos generales, se puede medir a través de la relación entre la media y la mediana. Si la media y la mediana son iguales, se dice que la distribución es simétrica; si la media es mayor que la mediana se dice que la distribución es sesgada a la derecha o que tiene sesgo positivo; si la media es menor que la mediana, se dice que la distribución es sesgada a la izquierda o que tienen sesgo negativo.

En la Tabla 1.10 se muestra un resumen general de la variable longitud de las cintas adhesivas.

Tabla 1.10: Estadísticas descriptivas de longitud de las cintas adhesivas.

Variable	Media	MediaRec	Desv.Est.	Varianza	CoefVar	Mnimo	Q1
longitud	8.264	8.082	3.641	13.259	44.06	2.37	5.39
Variable	Mediana	Q3	Mximo	Rango	RIC	Asimetra	Curtosis
longitud	7.373	10.71	19.355	16.986	5.319	0.68	0.03

Dado que el valor del coeficiente de asimetría es positivo (0.68) se puede afirmar que la distribución de los datos está sesgada a la derecha. Esta tabla permite también afirmar que los datos son heterogéneos, utilizando el criterio mencionado en los indicadores de dispersión. Esto se deduce del valor alto en el coeficiente de variación (44.06 %).

### 1.4.4.3. La regla empírica

Cuando la distribución de los datos tiene forma acampanada, los valores tienden a agruparse alrededor de la media y la mediana. En las distribuciones de esta forma, la regla empírica permite examinar la concentración de datos en intervalos medidos en desviaciones estándar por encima y por debajo de la media. Esta regla indica lo siguiente:

- Aproximadamente el 68 % de los valores se encuentran a una distancia de más o menos una desviación estándar de la media.
- Aproximadamente el 95 % de los valores se encuentran a una distancia de más o menos dos desviaciones estándar de la media.
- Aproximadamente el 99.7 % de los valores se encuentran a una distancia de más o menos tres desviaciones estándar de la media.

### Ejemplo.

Suponga que las longitudes de las varillas de hierro utilizadas en la elaboración de una viga de cemento cortadas en forma automática, tienen distribución simétrica con forma de campana, con un promedio de 8 pulgadas y una desviación estándar de 0.06 pulgadas. Para describir la longitud de estas varillas se puede utilizar la regla empírica de la siguiente manera:

$$1. \bar{x} \pm s = 8 \pm 0.06 = (7.94; 8.06)$$

$$2. \bar{x} \pm 2s = 8 \pm 2(0.06) = (7.88; 8.12)$$

$$3. \bar{x} \pm 3s = 8 \pm 3(0.06) = (7.82; 8.18)$$

Teniendo en cuenta los resultados anteriores, se puede afirmar que:

- Aproximadamente el 68 % de las varillas cortadas de esta forma tendrán longitud entre 7.94 y 8.06 pulgadas.
- Aproximadamente el 95 % de las varillas cortadas de esta forma tendrán longitud entre 7.88 y 8.12 pulgadas.
- Aproximadamente el 99.7 % de las varillas cortadas de esta forma tendrán longitud entre 7.82 y 8.18 pulgadas.

## 1.5. Resumen gráfico

Dependiendo del tipo de variables y su escala de medición, la representación gráfica puede ser:

### 1.5.1. Gráfico de torta

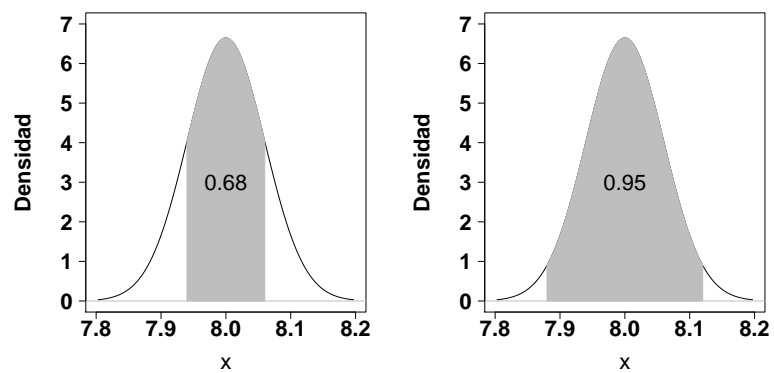
Este tipo de gráfico es pertinente en análisis de variables cualitativas con pocas alternativas. La Figura 1.4 muestra la representación gráfica para la el “tipo de defectos”, la cual tiene seis categorías.

En la Figura 1.4, se puede notar que la mayoría de artículos (76 %) no presenta ningún tipo de defecto y también muestra cómo se distribuye el 24 % restante entre los diferentes tipos de defectos, indicando que los defectos menor y mayor son los más comunes (11 % y 7 % respectivamente). Los defectos críticos, que son los más graves, se presentan en un 4 % de los artículos (3 % solo y un 1 % acompañado del defecto menor).

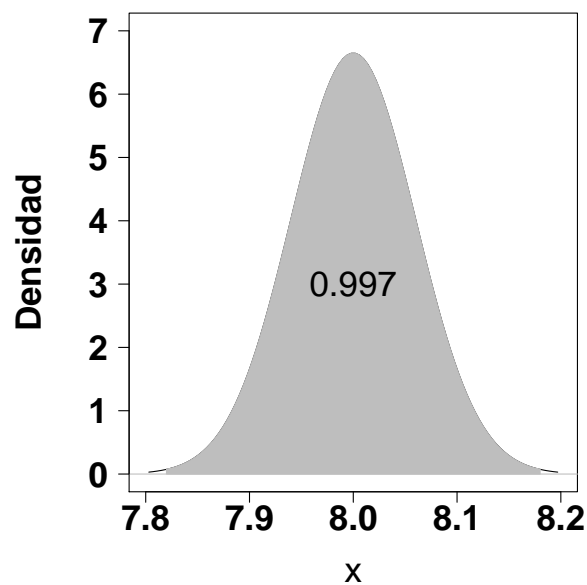
### 1.5.2. Diagrama de barras

De mayor uso en variables cualitativas, también es común verlo en variables cuantitativas discretas con pocos valores diferentes. La altura de la barra, indica la frecuencia (absoluta o relativa) y la base representa una categoría de la variable de estudio (o un valor de la variable cuantitativa discreta). Las barras deben ir separadas unas de otras y no implican un orden determinado (salvo el caso de las variables cuantitativas discretas). La Figura 1.5 muestra la distribución de la variable tipo de defecto.

El análisis de la información representada por este gráfico es similar al realizado con el diagrama de torta.



(a) Porcentaje aproximado de datos a una desviación de la media  
(b) Porcentaje aproximado de datos a dos desviaciones de la media



(c) Porcentaje aproximado de datos a tres desviaciones de la media

Figura 1.3: Regla Empírica aplicada a la variable longitud de las varillas en el ejemplo anterior

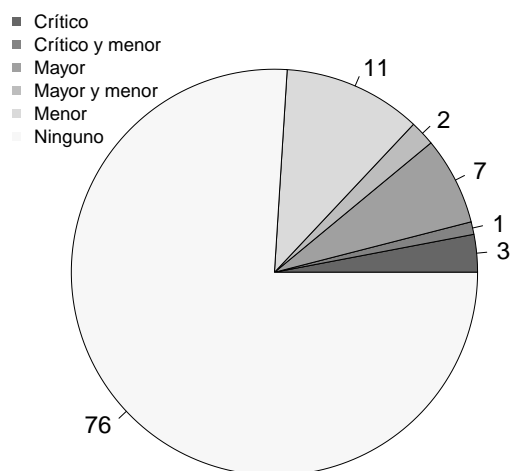


Figura 1.4: Distribución de tipo de defectos de los cojines.

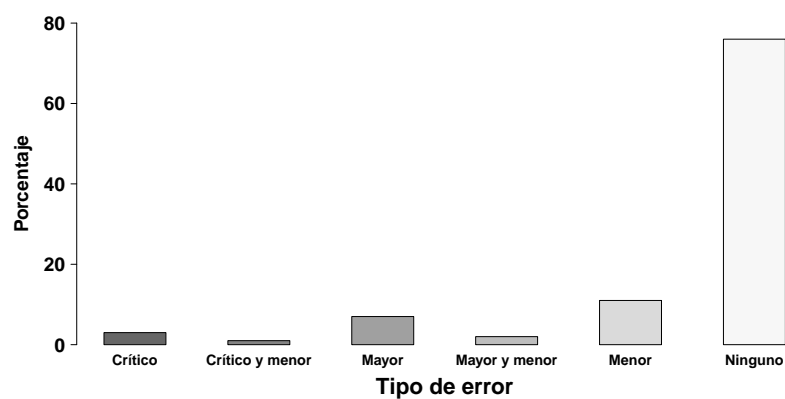


Figura 1.5: Distribución de los tipos de defectos de los cojines.

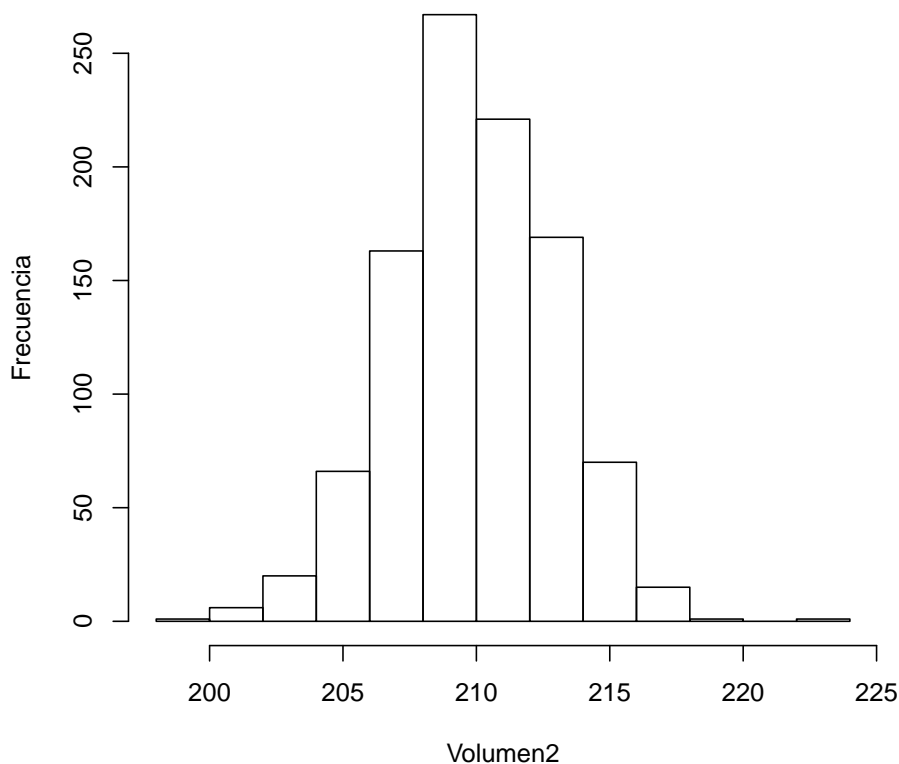


Figura 1.6: Distribución del volumen de aceite

### 1.5.3. Histograma

Es una gráfica de barras para datos cuantitativos agrupados en clases, donde no hay espacio entre las barras adyacentes por tratarse de variables continuas. La variable de interés se ubica en el eje horizontal ( $x$ ). La base del rectángulo (barra) representa al ancho de la clase; la altura del rectángulo (barra) representa la frecuencia, absoluta o relativa. Cuando las clases son de diferente ancho o amplitud, la altura del rectángulo debe ser siempre la densidad de frecuencia relativa, para poder interpretar adecuadamente la información contenida en el gráfico. Este diagrama permite identificar la forma como están distribuidos los datos. La Figura 1.6 muestra el histograma para la variable “volumen del aceite en presentación de 210 mililitros”

En la Figura 1.6 se puede apreciar como las frecuencias en cada uno de los intervalos permiten identificar la forma como están concentrados los datos, en este caso, la gráfica muestra una forma aproximadamente simétrica. La mayor concentración de datos para la variable volumen de aceite se presenta en las clases o intervalos de la parte central, entre 205 y 215 mililitros, donde está aproximadamente el 90 por ciento de los valores que toma la variable.



### 1.5.4. Polígono de frecuencia

Es similar al histograma en muchos aspectos, pero pretende dar una imagen aproximada de la curva definida por la distribución de los datos. Para construirlo se usan los mismos ejes que en el histograma; se señalan los puntos medios de cada intervalo en la parte superior del rectángulo y se conectan en forma consecutiva mediante un segmento, teniendo la precaución de marcar los puntos medios del intervalo que precede al primero y del que sigue al último. En la Figura 1.7 se presenta el polígono de frecuencia para la variable “volumen de aceite”.

Cuando se quiere comparar dos o más conjuntos de datos, los polígonos de frecuencia permiten mayor objetividad.

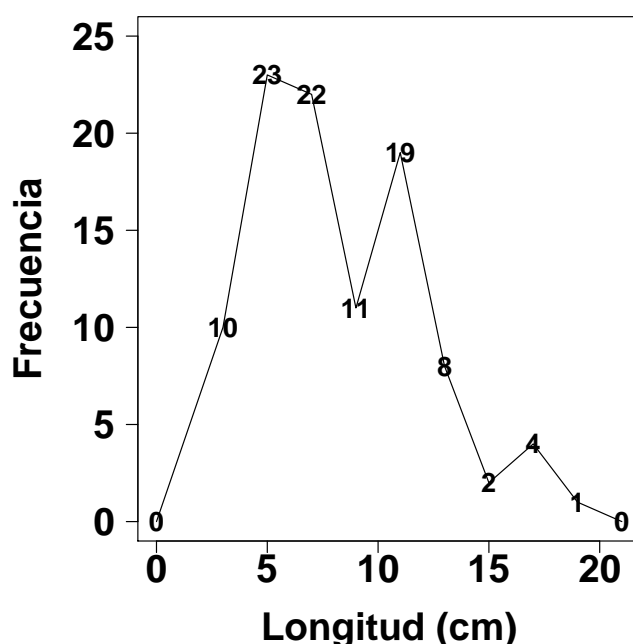


Figura 1.7: Distribución de la longitud de la cinta adhesiva.

En esta figura se puede notar que los valores que toma la variable están entre cero y veinte, la información se concentra al lado izquierdo, lo que sugiere una distribución asimétrica positiva ya que un gran porcentaje de los datos tienen valores que no superan los trece centímetros. La cola alargada del lado derecho, aunque tiene poca frecuencia, puede indicar que hay algunos productos que parecen estar fuera de control con respecto a esta variable.

### 1.5.5. Ojiva

Llamada también gráfico de frecuencia acumulado, muestra la variable de interés a lo largo del eje  $x$  y el porcentaje acumulado a lo largo del eje  $y$ . Para construirlo, se pueden usar las frecuencias relativas (o absolutas) acumuladas hasta los límites superiores de cada una de las clases indicadas en el histograma. Sirve para determinar qué porcentaje de los datos está por encima o por debajo de cierto valor, llamado percentil. En cada intervalo, la ojiva

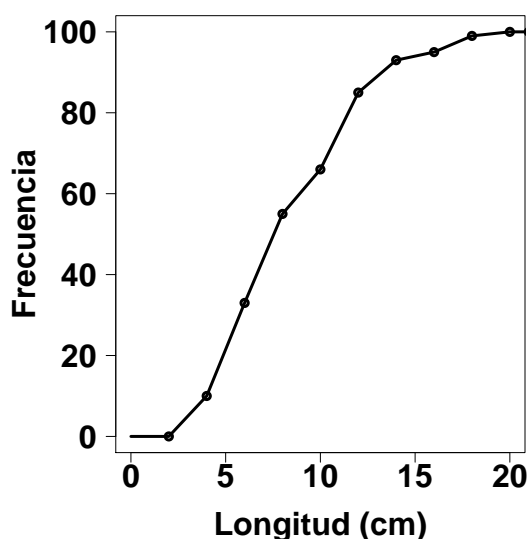


Figura 1.8: Distribución acumulada de la longitud de cintas adhesivas.

representa un segmento de recta, cuya pendiente es la densidad del intervalo respectivo. Cuando las clases son de diferente amplitud, la ojiva debe hacerse preferiblemente con la frecuencia relativa acumulada. En la Figura 1.8 se muestra la ojiva para la variable longitud de las cintas adhesivas. En este caso, la pendiente (concentración de los datos) es mayor en la segunda y tercera clase, seguido por la quinta clase.

### 1.5.6. Diagrama de tallo y hojas

Permite organizar los datos en intervalos o grupos, llamados tallos. Los valores dentro de cada grupo serán las hojas. El diagrama resultante permite ver cómo se distribuyen y donde están concentrados los datos, similar al histograma. En la Figura 1.9 aparece el diagrama de tallos y hojas para la variable peso de los cojines.

```

Tallo y hoja de Peso  N  = 100
Unidad de hoja = 0,010
 4      47  1234
 9      47  57999
16      48  0112334
22      48  568899
36      49  00000122222333
(16)    49  5666677777889999
48      50  01112234444
37      50  5556677777888999
21      51  222334444
12      51  5556899
 5      52  1
 4      52  56
 2      53  3
 1      53  5

```

Figura 1.9: Diagrama correspondiente a la distribución de la variable peso de los cojines

La primera columna representa la frecuencia acumulada y tanto de arriba-abajo como de abajo-arriba, finaliza en el intervalo que contiene la mediana, en el cual se muestra su frecuencia. La ventaja de este gráfico con respecto al histograma está en la originalidad

de los datos. En este diagrama no se usa punto medio, ya que se tiene el dato como tal. A partir de este gráfico se puede asegurar, entre otras cosas, de los datos se concentran entre 4.9 y 5.1 gramos

### 1.5.7. Diagrama de dispersión

Es un gráfico que involucra dos variables. Permite estudiar relaciones entre dos variables cuantitativas. Se dibuja un sistema de coordenadas cartesianas en el que se representan los valores que toman las dos variables para cada sujeto o unidad de análisis. La nube resultante de puntos permite evaluar si existe relación entre las dos variables y la naturaleza de la relación. Por ejemplo, si el analista de control de calidad de la empresa quiere saber si hay alguna relación entre las variables peso de los cojines y longitud de la cinta adhesiva, este tipo de gráfico es una buena alternativa. La Figura 1.10 muestra la relación entre las variables mencionadas.

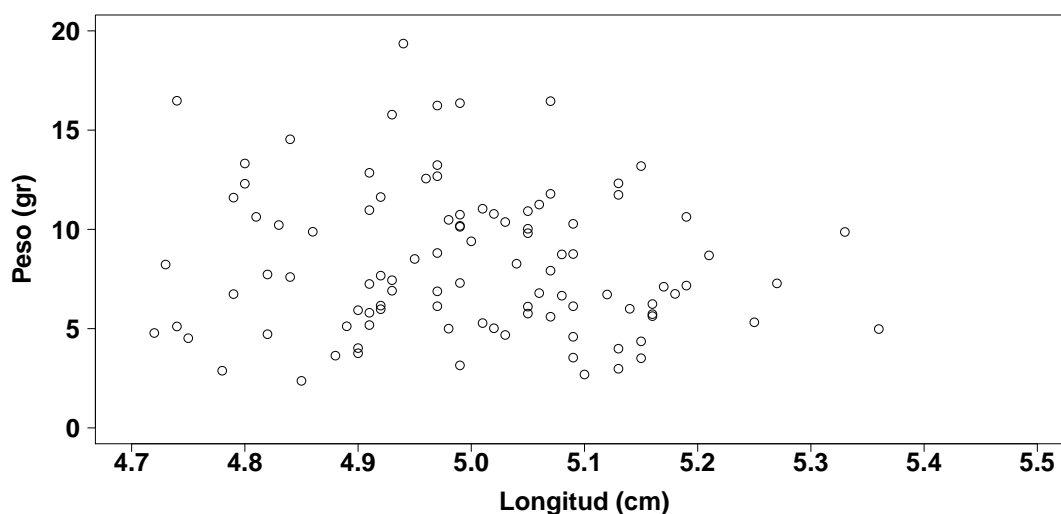


Figura 1.10: Relación entre las variable peso del cojín y longitud de la cinta

Este diagrama aparentemente no muestra ninguna relación entre las variables peso de los cojines y longitud de la cinta adhesiva.

En el capítulo cuatro se analiza el nivel de dependencia lineal entre dos variables y en el capítulo nueve se aborda la regresión lineal, mediante la cual se logra medir la relación entre variables en un modelo

### 1.5.8. Diagrama de caja y bigotes

Este diagrama ofrece una representación visual basada en el resumen de los tres cuartiles y de cualquier dato atípico que se presente en el grupo de datos. Permite visualizar cómo es la distribución de los datos. Es necesario definir algunos elementos. El rango intercuartílico ( $RIC$ ) se define como la diferencia entre el tercer y el primer cuartil ( $RIC = Q_3 - Q_1$ ).

Los datos atípicos se definen como puntos que son inusualmente grandes o pequeños. Para dibujar el diagrama de caja, cualquier punto que esté a más de 1.5 veces del rango intercuartílico por encima del tercer cuartil ó a más de 1.5 veces del rango intercuartílico por debajo del primer cuartil, se considera como datos atípicos.

Para construir el diagrama de caja se tiene en cuenta lo siguiente:

- Se halla la mediana, el primero y el tercer cuartil de los datos. Se trazan segmentos horizontales de igual longitud a la altura de cada uno de los valores anteriores y se dibujan segmentos verticales para completar la caja.
- Entre el dato más grande que no esté a más de 1.5 (*RIQ*) arriba del tercer cuartil y el dato más pequeño que no esté a más de 1.5 (*RIC*) debajo del primer cuartil; se dibujan segmentos verticales (bigotes) desde los puntos medios de los segmentos correspondientes al primer y tercer, hasta estos puntos.
- Puntos a más de 1.5 (*RIC*) arriba del tercer cuartil, o a más de 1.5 (*RIC*) por debajo del primer cuartil, se denominan datos atípicos. Cada dato atípico se dibuja por separado
- Si no hay datos atípicos, los bigotes correspondientes llegan hasta el menor y el mayor de los datos.

En las Figuras resumidas 1.11 aparecen los diagramas de caja y bigotes para las variables: peso de los cojines, volumen de aceite y longitud de las cintas adhesivas.

De las figuras anteriores se puede deducir que las variables "longitud de las cintas adhesivas y peso de los cojines", presentan un dato atípico, el cual es un valor muy grande comparado con el resto de sus valores. La variable "volumen de aceite", muestra un comportamiento aproximadamente simétrico

### 1.5.9. Diagrama de datos longitudinales o series temporales

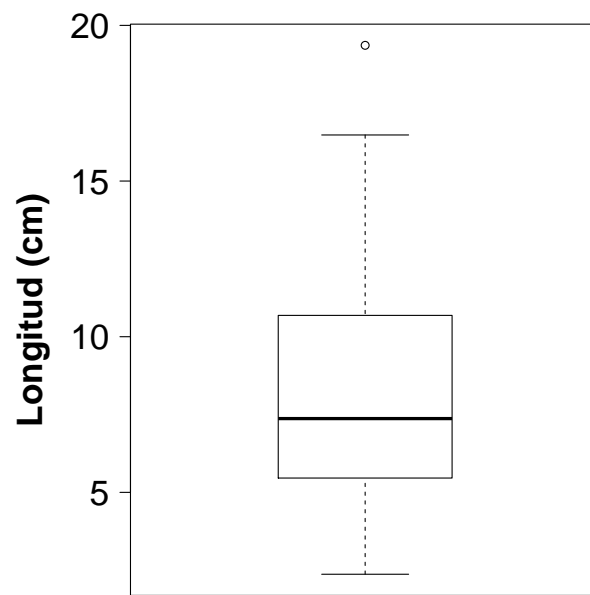
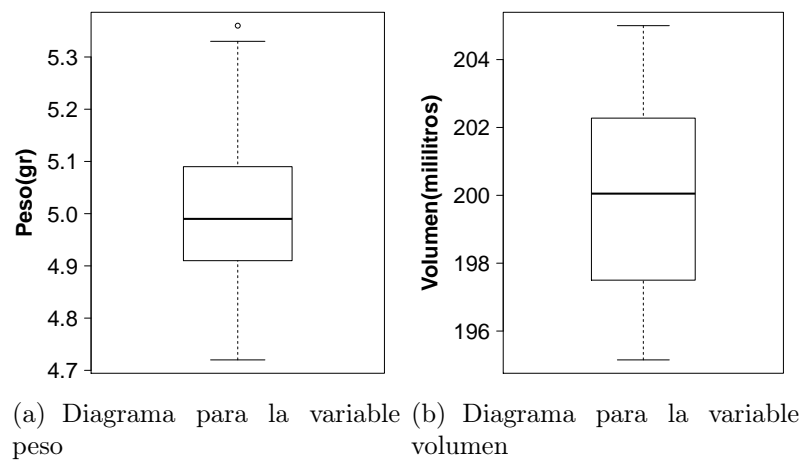
Una serie temporal es una secuencia de datos, observaciones o valores medidos en determinados momentos y ordenados cronológicamente. Los datos pueden estar espaciados a intervalos iguales o desiguales. En una serie temporal es importante la secuencia de las observaciones.

En una serie temporal, el tiempo se mide en el eje de las abscisas y la cantidad numérica que interesa, en el eje de las ordenadas. El gráfico de series temporales se obtiene uniendo los puntos contiguos en el tiempo por medio de segmentos de recta.

Las series de tiempo desempeñan un papel importante en el análisis requerido para el pronóstico de eventos futuros.

El tratamiento estadístico de las series de tiempo no es objeto de este libro, pero, es importante afirmar que existen modelos especiales que permiten su estudio.

En las figuras 1.12 y 1.13 se muestran las fluctuaciones del precio del dolar para diferentes intervalos de tiempo



(c) Diagrama para la variable longitud

Figura 1.11: Diagrama comparativo de las variables: peso, volumen y longitud

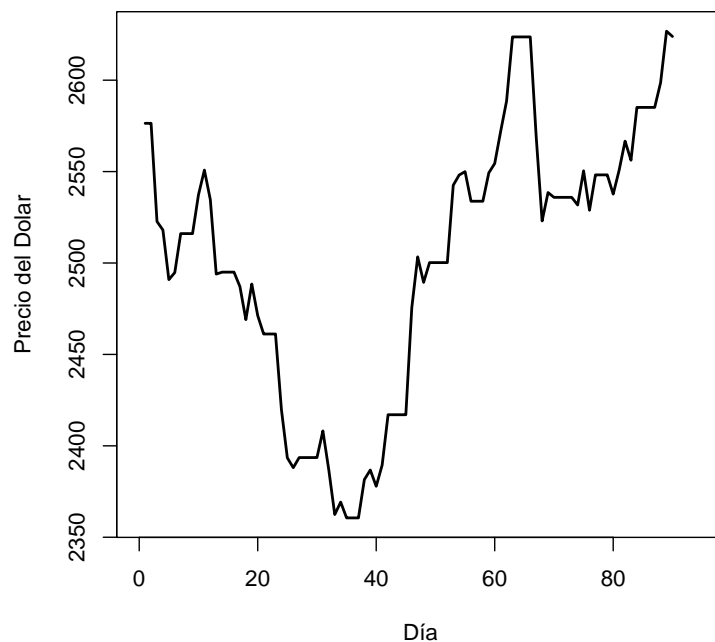


Figura 1.12: Fluctuación del dolar entre abril 05 y julio 02 de 2015

## 1.6. Códigos en R-Project

### 1.6.1. Variables Cualitativas

#### Leer base de Datos

```
datos=read.table("datos.txt",dec=";")
tabla=table(datos)
```

#### Diagrama de pastel

```
pie(tabla,labels=names("a","b"),col=c("red","blue"))
```

#### Diagrama de barras

```
barplot(tabla,names=c("a","b"),col=c("red","blue"))
```

### 1.6.2. Variables Cuantitativas

#### Resumen estadístico

```
summary(datos)
```

#### Media

```
mean(datos)
```

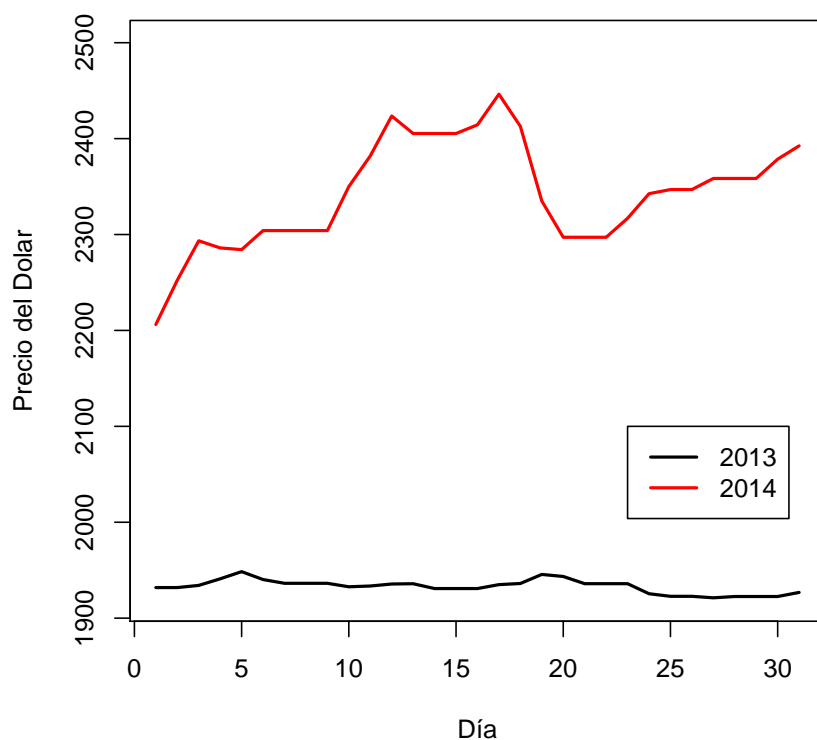


Figura 1.13: Comparación de la fluctuación del dolar para diciembre de 2013 y 2014

### Desviación estándar

`sd(datos)`

### Varianza

`var(datos)`

### Coefficiente de variación

`cv=mean(datos)/sd(datos)`

**Cuartiles** `quantile(datos,probs=c(0.25,0.5,0.75))`

### Histograma

`hist(datos,main="Nombre",col="yellow")`

### Diagrama de líneas

`plot(datos,type="l", col="red",main="Nombre",xlab="Eje x",ylab="Eje Y")`

### Diagrama de caja y bigotes

`boxplot(datos, col="blue", main="Nombre")`

**Ojiva**

```
ojiva.freq(h,type="l",xlab="Eje X",ylab="Eje Y",axes="")
```

**1.7. Ejercicios resueltos****1.7.1.**

En la siguiente tabla se reporta la información correspondiente a la concentración de sólidos suspendidos en agua de un río, (Tabla 1.9), en partes por millón, medidos semanalmente durante un largo período en una estación experimental. Se pretende utilizar el agua del río en un nuevo acueducto para la ciudad. Se debe elaborar de análisis descriptivo para la concentración de sólidos con el fin de validar la calidad del agua de este río

Tabla 1.11: Concentración de sólidos

55.8	60.9	37.0	91.3	65.8
45.9	39.1	35.5	56.0	44.6
83.2	40.0	31.7	36.7	62.3
75.3	71.4	65.2	52.6	58.2
60.7	77.1	59.1	49.5	69.3
42.3	33.8	60.6	76.0	69.0
71.7	61.2	61.5	47.2	74.5
47.3	94.6	56.3	30.0	68.2
48.0	61.8	78.8	39.8	65.0
69.8	64.9	27.1	87.1	66.3

**Solución**

Para el procesamiento de los datos se utilizará el software estadístico R.

La variable de interés es la concentración de sólidos (ppm).

Nótese que la variable concentración de sólidos está medida en decenas con precisión de tres cifras significativas

Por ser una variable continua, las frecuencias se generan por intervalos o clases que para este caso, resultaron 7 como se muestra en el Tabla 1.12. A partir de esta tabla se observa un comportamiento creciente en las frecuencias hasta el intervalo 4 y luego decreciente. Hay una concentración evidente de sólidos suspendidos hacia valores entre 57 y 67, representados por el valor medio de intervalo 62.0 partes por millón. Los intervalos a la izquierda se comportan de manera muy similar mientras que a la derecha si se observa variación en las frecuencias con tendencia a disminuir.

En el histograma, Figura 1.14, se puede corroborar la concentración de los sólidos hacia 62 y una leve asimetría positiva en tanto que el valor del sesgo es 0.0593, positivo y cercano a cero.

La Figura 1.15, diagrama de tallo y hojas, ofrece una distribución alternativa original para las frecuencias de los sólidos sin tener que recurrir al punto medio. Asigna tallo a la parte



Tabla 1.12: Distribución de frecuencias para sólidos suspendidos.

	Límite	Límite			Frecuencia	Frecuencia	Frecuencia
Clase	Inferior	Superior	Punto Medio	Frecuencia	Relativa	Acumulada	Rel. Acum.
1	27	37	32	7	0.14	7	0.14
2	37	47	42	6	0.12	13	0.26
3	47	57	52	8	0.16	21	0.42
4	57	67	62	14	0.28	35	0.7
5	67	77	72	9	0.18	44	0.88
6	77	87	82	3	0.06	47	0.94
7	87	97	92	3	0.06	50	1

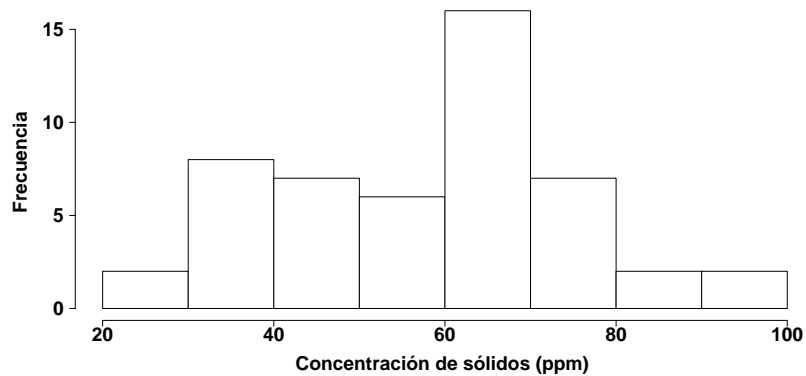


Figura 1.14: Sólidos suspendidos

Tabla 1.13: Resumen de Estadísticas

Recuento	50
Promedio	58.54
Desviación Estándar	16.4574
Coficiente de Variación	28.11 %
Mínimo	27.1
Máximo	94.6
Rango	67,5
Sesgo Estandarizado	0.0592781
Curtosis Estandarizada	-0.81852

de las decenas, desconociendo la parte decimal del dato y a las hojas, le asigna las unidades. El 27 por ejemplo lo separa en 2 para el árbol y 7 para la hoja. El segundo intervalo que va de 30-34 concentra los valores 30, 31 y 33; el tercer intervalo que va de 35-39 concentra los valores 35, 36, 37 y 39. Al girar el diagrama a la izquierda, se obtendría el histograma de frecuencias.

1	2 7
4	3 013
9	3 56799
12	4 024
17	4 57789
18	5 2
23	5 56689
(8)	6 00011124
19	6 55568999
11	7 114
8	7 5678
4	8 3
3	8 7
2	9 14

Figura 1.15: Distribución de los sólidos suspendidos

La Figura 1.16 de caja y bigotes agrupa la distribución en cuatro intervalos de 25 % cada uno. La caja inicia en  $Q_1$  y va hasta  $Q_3$ , alberga el 50 % de datos céntricos, divididos por la mediana (línea horizontal interna). Las extensiones van desde el mínimo a  $Q_1$  y de  $Q_3$  al máximo cuando no se presentan datos atípicos. No presentes en este caso, donde hay señal de una variación moderada en tanto que hay mayor concentración en el tercer grupo. También se aprecia una leve asimetría al presentarse diferencia entre la mediana y el promedio (señalado en la tabla 1.3).

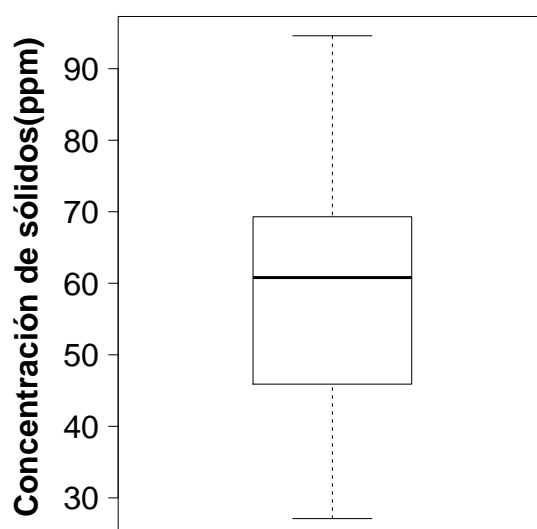


Figura 1.16: Concentración de sólidos suspendidos en agua

## 1.7.2.

Para estudiar el problema laboral que representa el ausentismo de los corteros de un ingenio azucarero, se realizó un estudio en el año 2005, sobre el ausentismo en este ingenio azucarero, buscando comparar el comportamiento específico de este con el comportamiento de ausentismo del gremio de ingenios de la región con relación a los 27 días de ausencia anual que presenta el gremio en la región. También conocer la distribución de estas ausencias con el fin de que departamento de recursos humanos tome los correctivos pertinentes. Los datos recaudados aparecen resumidos en la Tabla 1.14.

- Generar un reporte de análisis descriptivo a partir de la información obtenida.
- , Comparar que tan alejadas están las ausencias año con relación a los 27 días que presenta el gremio de la región

Tabla 1.14: Ausentismo de los corteros.

Días de ausencia	Frecuencia
0-10	20
10-20	12
20-35	10
35-50	8
50-70	6
70-90	3
90-105	1

**Solución:**

Los datos corresponden a la variable ausencia laboral en días por año y se mide en 60 corteros de caña de un ingenio. La variable es de tipo cuantitativa y está medida en escala de razón.

$Q_1$ ,  $Q_3$  y la mediana representan los cuartiles. Como se cuenta con datos agrupados por frecuencias, los cuartiles se aproximan mediante la expresión de la frecuencia acumulada propuesta en la Tabla 1.6.

$$P_k = L_{i-1} + \left( \frac{k}{100} - F(L_{i-1}) \right) C_i / f_i$$

Para esto, se tiene en cuenta que el análisis se hace con 60 corteros. El punto medio, representa al subconjunto de valores en cada intervalo. La frecuencia relativa y la relativa acumulada son usadas para el cálculo de los percentiles. El percentil 25 o  $Q_1$  se encuentra en el primer intervalo 0-10 días de ausencia. Puede calcularse a partir de una regla de tres o usando la fórmula anterior con los datos de la Tabla 1.15 y obtener:  $Q_1 = 0 + (0.25 - 0)10/0.3333 = 7.51$ . Indica esto que el 25 % de los corteros faltan menos de 8 días. De igual manera,  $Q_3$ , que se ubica en el cuarto intervalo (porque hasta el tercer intervalo hay 70 por ciento de los datos menor que 0.75 y hasta el cuarto hay un 83 por ciento, mayor que 0.75) entre 35 y 50 días.  $Q_3 = 35 + (0.75 - 42/60)15/0.1333 = 40.63$  Indica que el 75 % de los corteros falta

Tabla 1.15: Ausentismo en los corteros

Días de ausencia al año	Frecuencia Absoluta N	punto medio	Frecuencia Relativa f	Frecuencia Acumulada	
				Absoluta	Relativa
0-10	20	5	0.3333	20	0.3333
10-20	12	15	0.2000	32	0.5333
20-35	10	27.5	0.1667	42	0.7000
35-50	8	42.5	0.1333	50	0.8333
50-70	6	60	0.1000	56	0.9333
70-90	3	80	0.0500	59	0.9833
90-105	1	97.5	0.0167	60	1.0000

aproximadamente máximo 41 días. También puede decirse que el 25 % de los corteros en estudio falta aproximadamente al menos 41 días.

Así mismo del modelo se obtiene el segundo cuartil, el cual está ubicado en el segundo intervalo, entre 10 y 20 días.  $Q_2 = 10 + (0.5 - 0.3333)10/0.2 = 18.34$  lo que indica que el 50 % de los corteros está faltando hasta 18 días al año aproximadamente.

Al acumular los corteros que faltan hasta 50 días, se obtiene que 50 corresponde al límite superior del cuarto intervalo (35-50) y representan el 83.33 % de los datos  $(50/60)100\% = 83.33\%$ . Estos quiere decir que el 83.33 % de los corteros falta 50 días o menos.

Se cuenta con 7 intervalos a los que se debe representar con el punto medio para usar el modelo que aproxime el promedio con datos agrupados, el cual se obtiene mediante la expresión  $\bar{x} = \frac{\sum x'_i n_i}{n}$  planteada en la Tabla 1.6. Así,  $\bar{x} = (5)(20) + (15)(12) + (27.5)(10) + (42.5)(8) + (60)(6) + (80)(3) + (97.5)(1)/60 = 26.54..$

Esto indica que cada cortero está faltando alrededor 26.5 días al año. Como este valor es muy similar al reportado por la asociación de ingenios, significa que en la fábrica se está presentando aproximadamente el mismo indicador regional.

Si la empresa quisiera incentivar a los corteros que falten 3 días o menos. Contrario al proceso de hallar el percentil cuando se conoce el acumulado, este valor se puede obtener mediante la expresión  $F(x) = F(L_{i-1}) + \frac{f_i}{C_i}(x - L_{i-1})$ , planteada en la Tabla 1.6. En este caso particular, el valor que se quiere hallar es  $F(3)$ , el cual debería encontrarse entre 0 % y 3.3 % puesto que el valor 3 de la variable  $X$ , se encuentra en el primer intervalo. De acuerdo la Tabla 1.15, este valor se puede encontrar de la siguiente manera:  $F(3) = 0 + (3 - 0)0.333/10 = 0.09999$ , correspondiente aproximadamente a 10 %. Lo anterior significa que la empresa premiará al 10 % de corteros.

## 1.7.3.

Una empresa distribuidora de cemento tiene cuatro locales en una ciudad. La información resumida en la Tabla 1.16 y Figura 1.17, realizadas en R-Project, corresponde a las ventas diarias de bultos de cemento en una muestra aleatoria de diez días, procedente de cada uno de los cuatro locales.

Tabla 1.16: Estadísticas descriptivas: Ventas

Local	Media	Mínimo	Q1	Mediana	Q3	Máximo	Rango	RIC
1	1,010	600	775	1,050	1,225	1,400	800	450
2	1,020	100	275	1,050	1,725	1,900	1,800	1,450
3	1,840	200	1,500	2,100	2,500	2,600	2,400	1,000
4	1,250	800	975	1,050	1,475	2,200	1,400	500

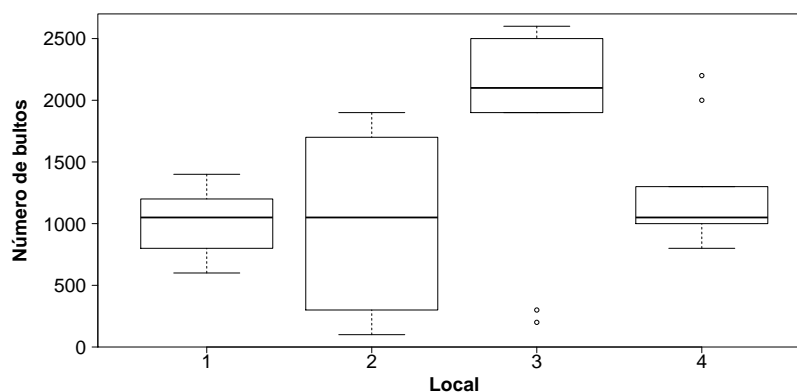


Figura 1.17: Comparación de las ventas de bultos de cemento por local

Al observar las estadísticas descriptivas y el diagrama de caja y bigotes, se puede decir que en los locales 1, 2 y 4, tienen aproximadamente las mismas ventas medianas de 1,050 bultos diarios. Los locales 1 y 2 tienen aproximadamente el mismo promedio de bultos vendidos. Al examinar los rangos y rangos intercuartílicos, se puede apreciar que las ventas del local 2 están más dispersas que las del local 1. El local 3 tiene las ventas medias y medianas más grandes que el resto de locales (1,840 y 2,100 respectivamente), pero, también la mayor dispersión (mayor rango).

Se puede observar que la distribución de las ventas del local 3 está sesgada hacia la izquierda, lo cual refleja la presencia de días con ventas muy dispersas en el 25 por ciento inferior. De igual forma, la distribución de las ventas del local 4 está sesgada hacia la derecha, lo que refleja la presencia de algunos días con ventas muy dispersas a partir del tercer cuartil.

Para la empresa es muy importante conocer el comportamiento de las ventas tanto en un local determinado como en los cuatro locales, por esto los indicadores de posición y dispersión son de mucha utilidad.

**1.7.4.**

Tomado del texto de la profesora Luz Elena Vinasco, de la facultad de ingeniería

Una empresa productora de panela cuyo producto se realiza solo con la obtención de mieles de caña de azúcar realiza un plan de mejoramiento del proceso productivo de la panela.

La metodología utilizada para desarrollar la evaluación y diagnóstico de la empresa es la siguiente:

- Recolectar la información en la planta constituida por hechos, toma de datos, opiniones, apreciaciones y consultas sobre el proceso y funcionamiento.
- Toma de índices, tiempos, variables de control, para la detección de los puntos críticos
- Analizar la información identificando los puntos críticos o situaciones que requieren una acción de mejoramiento

Cuando las mieles alcanzan el punto de panela se lleva en los bates hasta el área de moldeo en la cual están ubicadas las gaveras o conjunto de moldes donde se termina de enfriar y solidificar hasta alcanzar su forma definitiva de panela, también se encuentran los "pesadores", operarios encargados de dar forma a la panela con los cocos de madera.

Las gaveras se encuentran ubicadas sobre mesas de cemento cubiertas con paños, estas gaveras se humedecen antes de depositar en ellas las mieles para evitar que la panela se pegue al molde. Con la ayuda de palas de madera, el operario distribuye la panela en cada uno de los moldes de la gavera.

Estos cocos y las gaveras son un centro de contaminación debido a que humedecen con agua no tratada, ocasionando crecimiento de hongos que se ve reflejado con una capa mucosa adherida en la madera de los cocos y moldes, así como también problemas de contaminación en los operarios o pasadores de panela.

Es evidente que en estas condiciones la panela se contamine y por ende su calidad se deteriore. Por otro lado, se encuentra la dificultad de estandarizar el peso del producto, debido a que este depende de las herramientas utilizadas y del operario.

A continuación se mencionan los pasos involucrados en el proceso de fabricación de la panela de la empresa donde se realizó el estudio.

- Etapa de corte
- Etapa de molienda
- Etapa de limpieza de los jugos
- Etapa de evaporación
- Etapa de punteo

- Etapa de batido
- Etapa de moldeo
- Etapa de empaque
- Etapa de almacenamiento y distribución

El objetivo general del estudio es plantear propuestas de mejora para disminuir la variación de los pesos de la panela corriente

Entre los objetivos específicos de tiene:

- Ordenar los datos recolectados y analizarlos aplicando herramientas estadísticas para mostrar que implican los defectos para la empresa
- Diseñar una nueva herramienta de moldeo que reduzca la variación de los pesos
- Implementar propuestas de mejora al problema identificado
- Cuantificar el impacto de las propuestas implementadas

La empresa no cuenta con un sistema adecuado para la elaboración del moldeo de la panela. El problema consiste en que el encargado de esta tarea no posee las herramientas adecuadas para verificar la cantidad de panela que adquiere por medio de los moldes de madera, produciendo unidades de panela con pesos inadecuados y de gran variabilidad. Esto afecta de manera evidente el rendimiento productivo de la empresa ya que el exceso de peso y la falta de peso, genera pérdidas para la empresa.

Teniendo en cuenta que la empresa no lleva el control de calidad requerida en la línea de panela en presentación "panela corriente"; el primer paso es registrar el peso de cada panela. Se toma una muestra de los pesos de cada panela producida por día hábil para el mes de febrero de 2011 ( $n=210$  panelas). La información se presenta en el archivo XXX1

Teniendo en cuenta que el peso promedio conque debe salir la panela es 502.5 gramos y que el resultado del procesamiento de la información indica un peso promedio de 520 gramos, ver Tabla 1.17, se puede afirmar que las panelas están saliendo con un sobrepeso promedio de 18 gramos aproximadamente.

Las Figuras 1.18 y 1.19 muestran que los pesos de las panelas no están cumpliendo con las especificaciones de la empresa, debido a que el primer cuartil es 502 gramos, lo que indica que el 75 % de la producción de panela está saliendo por encima del promedio establecido; además, algunos pesos de panelas son muy superiores al resto de la información, es decir, que se consideran datos atípicos .

Los efectos del sobrepeso y la variación en el peso de la panela corriente genera pérdida del producto, lo cual se va a ver reflejado en los ingresos de la empresa. Dado que la empresa ha establecido como límites de especificaciones: inferior de 495 gramos y superior de 510 gramos, en la Figura 1.18, se puede observar que existe un gran porcentajes de pesos de las panelas que están por fuera de estos límites. Debido a lo anterior, se recomienda

Tabla 1.17: Estadísticas peso de panela antes

	Pesoantes
Recuento	210
Promedio	520,043
Desviación Estándar	23,1804
Coefficiente de Variación	4,46 %
Mínimo	460
Máximo	580
Rango	120
Sesgo Estandarizado	1,88578
Curtosis Estandarizada	0,297571

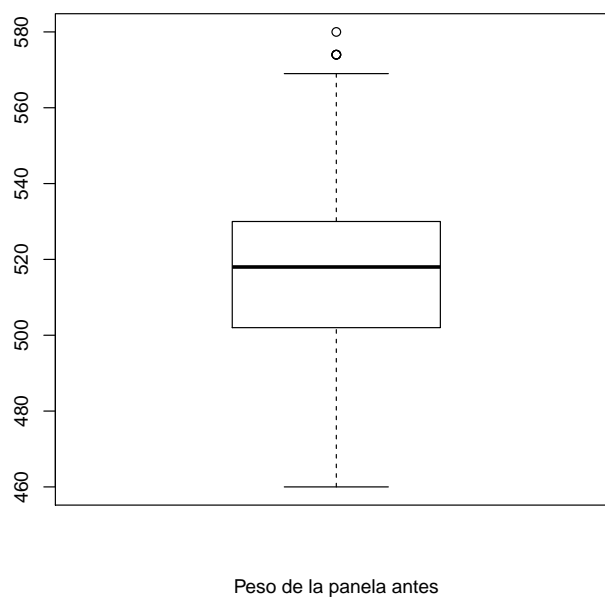


Figura 1.18: Distribución del peso de la panela



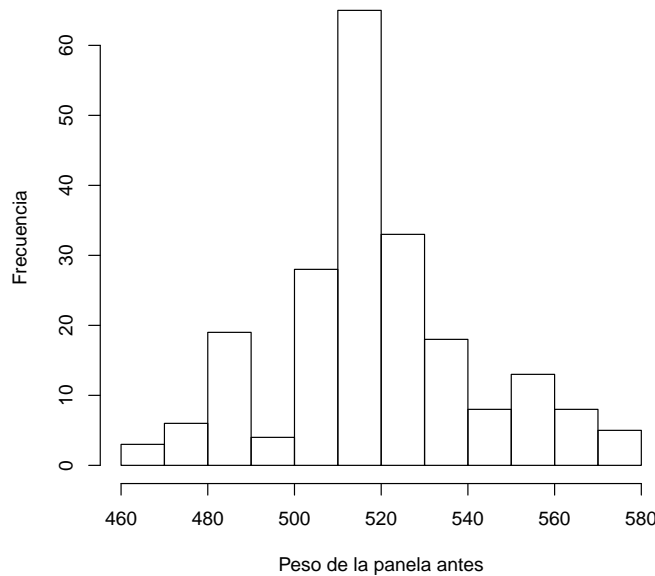


Figura 1.19: Histograma del peso de la panela antes

implementar una propuesta de mejora en el área de moldeado de la panela corriente.

Al identificar las causas del problema e implementar mejoras en el proceso de producción de la panela con presentación "panela corriente" se pretende disminuir la dispersión y cumplir con el objetivo del peso propuesto por la empresa.

Con el fin de identificar las causas del problema, mediante una lluvia de ideas se intenta contestar a la pregunta: ¿Porqué en el proceso productivo de la empresa se encuentra variabilidad en los pesos de las panelas?

Después de identificar las posibles causas, las propuestas a implementar con respecto a los problemas son:

- Capacitaciones a los operarios en cuanto a la seguridad industrial, concientizándolos de los riesgos existentes en la planta
- Dar incentivos a los trabajadores sin que llegue a incurrir en altos costos para la empresa, con el fin de aumentar la productividad
- Realizar acciones preventivas y no correctivas
- Implementar pausas activas para evitar la fatiga del operario

Luego de haber estudiado previamente toda la información adquirida del trapiche, se determinó que el problema más representativo fue en el área donde se moldea la panela; se realiza esta propuesta con el fin de brindarle la forma y su medida correcta a la panela

antes de sacarla al mercado, utilizando moldes con unas especificaciones diferentes

Para la implementación de estos moldes, se trabajará con el operario con mayor experiencia en el área de moldeo, el cual recibirá una capacitación previa con el fin de estandarizar los movimientos a realizar para obtener un producto con las especificaciones deseadas por la empresa.

Después de haber realizado la capacitación, se recolectaron datos donde los resultados obtenidos fueron satisfactorios, motivando a la empresa a realizar una inversión en la compra de los moldes para el uso exclusivo del operario experto.

La acogida de la propuesta por la empresa ha sido satisfactoria, lo cual ha hecho que el aspecto físico de la panela se vea mucho mejor, su color es homogéneo y su forma es más elegante.

Para medir el impacto en el mejoramiento de la calidad en la línea de producción de la panela, se tomó información de los pesos de las panelas durante el mes de abril (210 panelas). La información se encuentra en el archivo xxx2

En la Tabla 1.18 se observa la disminución de la dispersión de los pesos de la panela después de realizada la implementación en la línea de producción; antes el peso promedio era de 520 gramos y después el peso bajó a 511 gramos, acercándose más a la meta establecida por la empresa que es de 502 gramos, en cuanto la variación del peso, antes era de 23 gramos y después de implementadas las propuestas de mejora bajó a 15 gramos, lo que significa que se ha logrado estandarizar un poco más el proceso, ya que los pesos son un poco más estables, aunque aún no se ha llegado al peso deseado.

Tabla 1.18: Estadísticas peso de panela antes y después

	Pesoantes	Pesodespues
Recuento	210	210
Promedio	520,043	511,029
Desviación Estándar	23,1804	15,167
Coefficiente de Variación	4,46 %	2,97 %
Mínimo	460	466
Máximo	580	545
Rango	120	79
Sesgo Estandarizado	1,88578	-1,80246
Curtosis Estandarizada	0,297571	0,107948

En la Figura 1.20 se ilustra el gráfico de caja y bigotes, para el peso de las panelas producidas antes y después de realizar la implementación de las propuestas, en donde se puede observar la disminución tanto en el promedio como en la dispersión del peso.

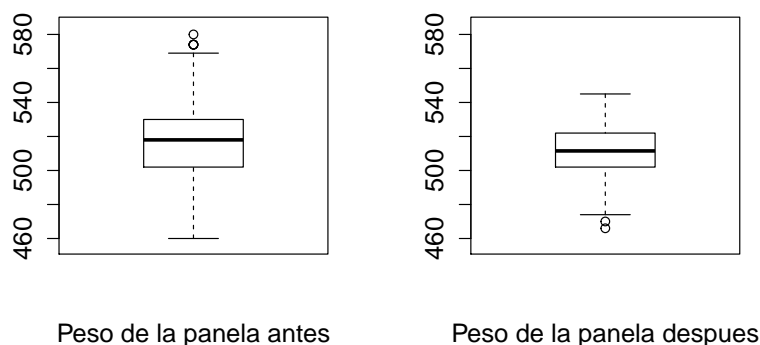


Figura 1.20: Diagrama de caja comparativo del peso de la panela

En los pesos después de la implementación se observa que se presentaron algunas panelas muy por debajo del límite de especificación establecido, lo que alerta para identificar las causas y solucionar este problema.

Entre las conclusiones de esta caso se puede mencionar lo siguiente:

- Gracias a los moldes se logró obtener una forma estándar de la panela, evitando tener desperdicios en cuanto al empaque
- A partir de los análisis estadísticos, se concluye que la implementación de los moldes realizada en el trapiche tuvo un impacto de mejora notable en cuanto al peso mas no al óptimo, sin embargo es necesario para la empresa seguir trabajando en el método hasta lograr el pedo ideal de la panela.

## 1.8. Problemas Propuestos

### 1.8.1.

El departamento de control de calidad de una empresa productora de jugos decide estudiar el comportamiento de la máquina que etiqueta las botellas de jugos que van a salir al mercado, en cuanto a la velocidad y variabilidad del número de botellas etiquetadas diariamente, con el propósito de validar la posibilidad de cambiar la máquina. Decide hacer seguimiento al número de botellas etiquetadas por día durante un período de ocho semanas. Los datos registrados se presentan en la Tabla 1.19.

Presentar un análisis estadístico que le ayude al departamento de control de calidad a tomar la decisión.

### 1.8.2.

Se realizó un experimento con el fin de comparar el contenido del hierro en las barras metálicas para un proceso a partir de un mecanismo manual y otro mecánico. Los datos

Tabla 1.19: Número de botellas etiquetadas por día

7532	9121	7396	8011	8736	7470	8643	7883
8457	7569	8254	8115	7693	8019	7237	6897
6275	6578	7453	7457	8457	6581	7593	8457
5944	6991	7032	8745	6337	6825	8240	5985
8963	7932	6854	7253	7352	7318	6953	6429
6501	6549	8337	6485	8119	6584	7495	6867

de la Tabla 1.20 son mediciones del contenido de hierro realizadas en 16 barras metálicas seleccionadas al azar de un lote.

- Para cada método de medición, obtenga la medida de tendencia central que mejor resume el conjunto de datos.
- Cuál método de medición genera menor dispersión?
- Compare los métodos a través del diagrama de caja y bigotes. ¿Se presentan datos atípicos para alguno de los métodos? Justificar su respuesta.

Tabla 1.20: Cantidad de hierro

Barra número	mecánico	manual
1	63.22	63.27
2	61.08	64.20
3	62.66	62.94
4	62.87	62.17
5	63.01	61.89
6	61.75	63.92
7	62.10	64.34
8	63.57	63.64
9	63.22	62.10
10	61.68	63.64
11	62.08	62.87
12	62.87	62.30
13	64.43	63.50
14	62.87	62.03
15	62.55	65.10
16	62.15	62.73

### 1.8.3.

En la inducción a un nuevo Administrador, se le da la oportunidad de que haga un diagnóstico de la empresa en la cual aspira trabajar. Como cuenta con poco tiempo, cree que ganará respeto en el periodo de prueba si presenta en su primer informe un análisis estadístico sobre la información que se tiene sobre el contenido de aceite industrial en presentación de mil mililitros, por ser el producto menos comercializado y posteriormente

proponer una estrategia de mercadeo para mejorar su posicionamiento en el mercado. Retomando la estadística, decide hacer análisis sobre el promedio y variación en el contenido empacado de este aceite y la proporción de clientes que están insatisfechos con este producto.

Consultando los archivos, detecta que la distribución en el contenido del aceite es acampanada en el empaque ya mencionado. La muestra aleatoria con la cual se cuenta presenta los siguientes valores.

1,003 990 1,000 991 997 1,001 1,003 998 996 996 995 1,005 1,004 1,000 1,000 999 1,005 995 992 994

Los aspectos sobre los cuales decide presentar su informe están enfocados a los siguientes puntos.

- El riesgo de que la empresa sea demandada por incumplimiento en el contenido promedio del producto.
- Dada que la muestra fue clasificada según el turno en que se empacó, determinar si existe diferencia tanto en promedio como en variabilidad entre los turnos.
- Respecto a la satisfacción de los clientes, ¿qué debería registrar en el informe, si en una encuesta a 60 clientes de este tipo de aceite, 12 dicen estar insatisfechos con el producto?

Tabla 1.21: Contenido de aceite en presentaciones de mil mililitros

<b>Diurno</b>	997	994	990	1,000	992	1,005	999	996	1,000	1,004	1,001
<b>Nocturno</b>	1,000	1,005	1,003	996	998	1,003	991	995	995		

Tabla 1.22: Resumen Estadístico para el contenido de aceite

	Turno=diurno	Turno=nocturno
Recuento	11	9
Promedio	998.0	998.444
Desviación Estándar	47.3286	46.398
Coficiente de Variación	0.474235 %	0.464703 %
Mínimo	990.0	991.0
Máximo	1005.0	1005.0
Rango	15.0	14.0
Sesgo Estandarizado	-0.309075	-0.067828
Curtosis Estandarizada	-0.476827	-0.639464

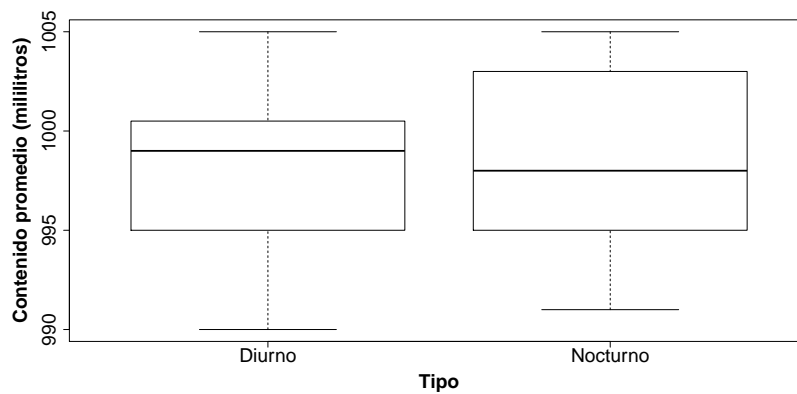


Figura 1.21: Contenido de aceite empacado por turno

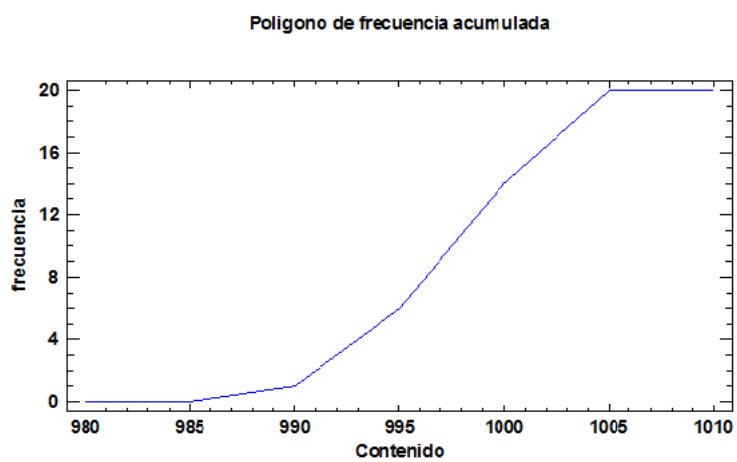
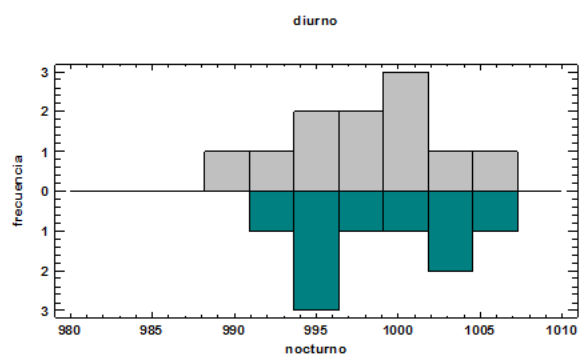


Tabla 1.23: Ventas realizadas sin promoción para una tienda

184	157	187	258	125	106	451	379	184	301
120	135	139	106	180	150	623	339	554	179

Tabla 1.24: Ventas realizadas con promoción para una tienda

755	357	810	101	660	725	303	201	785	420
357	260	129	560	135	432	808	184	156	727

#### 1.8.4.

En un estudio de mercados para una tienda, consideran que la promoción "descuentos del 20 % en toda la tienda", es una herramienta que ayuda a incrementar las ventas. Los investigadores examinaron la efectividad de esta promoción y se tomó información del número de ventas diarias durante veinte días antes y después de la promoción. La información se presenta en las Tablas 1.23 y 1.24

- Analice la efectividad de las promociones que se realizaron en la tienda.

#### 1.8.5.

Se seleccionó una muestra de 20 botellas de vidrio de determinado tipo, y se midió la resistencia de cada una a la presión interna. Tomar en cuenta la siguiente información parcial de la muestra: Mediana= 202.2, primer cuartil: 196.0 y tercer cuartil: 216.8.

Tres observaciones mínimas: 125.8, 188.1, 193.7

Tres observaciones máximas: 221.3, 230.5, 250.2

- Verificar si hay datos atípicos en la muestra.
- Construir el diagrama adecuado que le permita justificar la respuesta del punto anterior.
- Establecer y argumentar el tipo de distribución que tienen estos datos

#### 1.8.6.

La madera que se emplea en la construcción de edificios debe monitorearse en cuanto a su resistencia. Datos para la resistencia de piezas de madera, utilizada para pisos del edificio, en libras por pulgada cuadrada, se proporcionan en la Tabla 1.25.

Tabla 1.25: Datos de madera empleada en la construcción de edificios

1325	1419	1490	1633	1645	1655	1710	1712	1725	1727	1745	1828	1840	1856	1859
1867	1889	1899	1943	1954	1976	2046	2061	2104	2168	2199	2276	2326	2403	2983

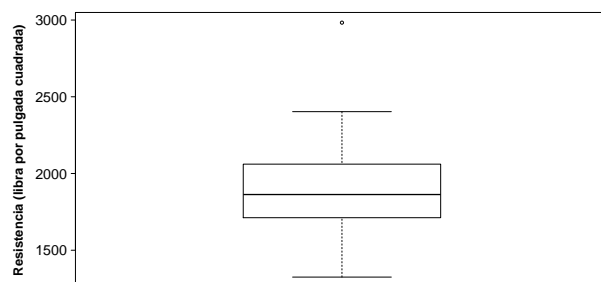
Verificar utilizando el programa estadístico R-Project si los siguientes resúmenes son correctos

Tabla 1.26: Resumen estadístico para resistencia

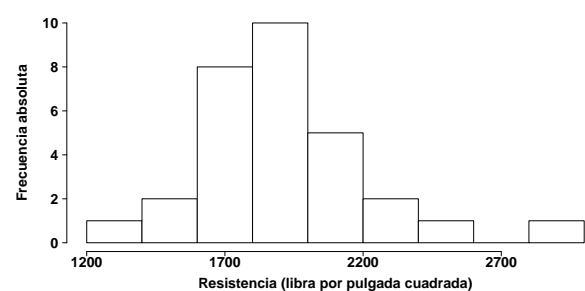
Recuento	30
Promedio	1908.77
Desviación Estándar	327.115
Coefficiente de Variación	171375 %
Mínimo	1325.0
Máximo	2983.0
Rango	1658.0
Sesgo Estandarizado	250.083
Curtosis Estandarizada	322.368

Tabla 1.27: Percentiles correspondientes a la resistencia de la madera

Percentiles	
1.0 %	1325.0
5.0 %	1419.0
10.0 %	1561.5
25.0 %	1712.0
50.0 %	1863.0
75.0 %	2061.0
90.0 %	2301.0
95.0 %	2403.0
99.0 %	2983.0



(a) Resistencia de las piezas de madera



(b) Resistencia de las piezas de madera

- Realizar un análisis numérico a partir de la información de la muestra
- A partir de los gráficos complementar el análisis realizado para la parte numérica.

### 1.8.7.

Una característica de calidad importante en el proceso de llenado de bolsas de café instantáneo es el peso de cada una de las bolsas. La empresa no puede correr el riesgo de empacar bolsas por debajo del peso anunciado en la etiqueta porque puede incurrir en multas, además, si las bolsas tienen menos contenido de lo que deberían tener, posiblemente los clientes no puedan preparar un café tan fuerte como lo deseen. Si por el contrario, las bolsas se empacan muy por encima del peso que debe ser, se estará regalando el producto e



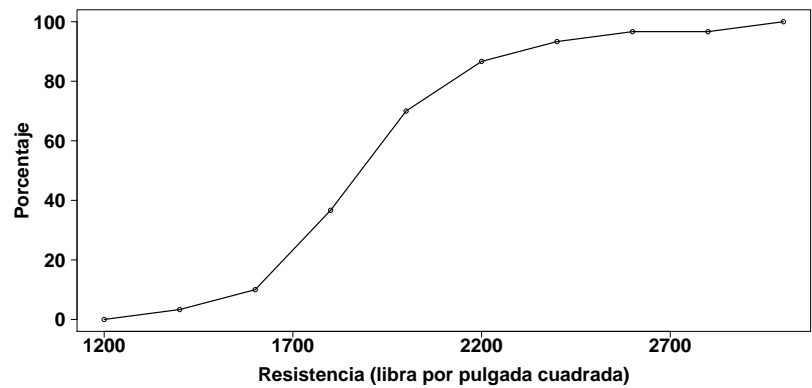


Figura 1.22: Resistencia de las piezas de madera

Tabla 1.28: Peso de las bolsas de café (gramos)

6.20	5.22	6.74	7.78	7.70	8.23	4.32	6.27	7.60	5.41
5.81	4.81	4.65	5.52	5.73	4.38	5.93	6.10	6.63	6.13
6.17	6.13	7.84	6.41	6.31	5.99	8.47	7.37	8.88	5.85
8.16	4.89	7.04	7.40	8.42	6.42	5.98	7.18	6.12	7.26
5.06	5.65	4.98	6.14	6.47	6.53	6.18	8.69	4.76	5.76

incurriendo en sobrecostos para la empresa. La etiqueta del peso en la bolsa indica que en promedio, debe haber 6.5 gramos de café. Un objetivo de calidad es que la empresa llene las bolsas con la cantidad etiquetada. La siguiente tabla proporciona el peso en gramos de una muestra de cincuenta bolsas llenadas por la máquina en una hora.

- Obtener los gráficos e indicadores adecuados para presentar un informe al departamento de control de calidad.
- Con los resultados encontrados en el punto anterior realizar el informe adecuado y plantear las sugerencias a tener en cuenta para resolver los problemas que pueda tener el empaclado.

1.8.8.

Un grupo de ingenieros civiles encargado de monitorear el caudal en ríos donde se genera electricidad. Las siguientes son las mediciones de caudal diario, en millones de galones (MGD), en el río Namekagon en el mes de mayo durante 47 años. Ver datos en el archivo XXX2

- Mostrar mediante el gráfico adecuado, la tendencia de los datos a través del tiempo
- Obtener e interpretar los cuartiles.
- Obtener los valores de los caudales a partir de los cuales se identifican los valores atípicos (si los hay). Si hay valores atípicos, indicar a que año corresponden
- Determinar los indicadores de posición central, dispersión y forma que permitan resumir los resultados del monitoreo.

## 1.8.9.

El gerente de una empresa constructora considera que sus tres grupos de trabajo no lo hacen con igual rapidez y además se queja de la permanente lentitud en sus grupos para presentar las propuestas de proyectos. Realiza mediciones del tiempo de retraso en la entrega de las propuestas de proyectos, resumidas en la Figura 1.23.

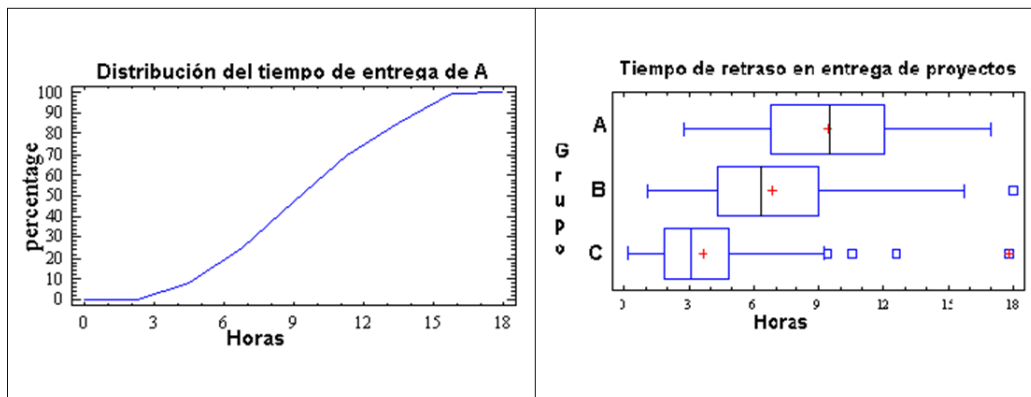


Figura 1.23: Tiempos de retraso de los grupos de trabajo en presentar propuestas de proyectos

- Comparar los tiempos medios y los cuartiles en las horas de retraso en cada grupo.
- Comparar la distribución de los tiempos de retraso en los proyectos de los tres grupos.

Para el grupo de trabajo A,

- ¿Por debajo de qué tiempo se encuentra el 40% de los proyectos entregados tardíamente?
- ¿Por encima de cuántas horas se encuentra el 75% de los proyectos?
- ¿El 40% de los proyectos se entregan en más de cuántas horas?
- Determinar y analice los percentiles 30, 60 y 85.
- ¿Qué porcentaje de proyectos se entregan en menos de 7 horas de retraso?
- Si la desviación estándar en A es 3.4, que porcentaje de proyectos se entregan a una desviación estándar de la media?
- Si en A se entregan 40 proyectos tardíos, cuántos de estos se entregan después de 15 horas?