

**DESARROLLO DE GUÍAS DE APRENDIZAJE SOBRE GRÁFICAS DE  
CONTROL MEDIANTE EL USO DEL SOFTWARE R**

**DUBAN CASTILLO CÁRDENAS**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
FACULTAD DE INGENIERÍA  
BOGOTÁ  
2020**

**DESARROLLO DE GUÍAS DE APRENDIZAJE SOBRE GRÁFICAS DE  
CONTROL MEDIANTE EL USO DEL SOFTWARE R**

**DUBAN CASTILLO CÁRDENAS**

**Trabajo de grado para optar al título de  
Ingeniero Industrial**

**Directora:  
Ing. Ruth Milena Suarez Castro**



**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES  
FACULTAD DE INGENIERÍA  
BOGOTÁ  
2020**

NOTA DE ACEPTACIÓN

---

---

---

---

---

---

---

---

Firma del presidente del jurado

---

Firma del jurado

---

Firma del jurado

Bogotá D.C día mes y fecha

## CONTENIDO

	pág.
GLOSARIO.....	9
INTRODUCCIÓN.....	12
1 PLANTEAMIENTO DEL PROBLEMA .....	13
2 JUSTIFICACIÓN.....	14
3 OBJETIVOS .....	15
3.1 Objetivo General.....	15
3.2 Objetivos Específicos.....	15
4 MARCO REFERENCIAL .....	16
4.1 Marco Teórico.....	16
4.1.1 Lenguaje R .....	16
4.1.1.1 El entorno $R$ .....	16
4.1.1.2 Funcionamiento de $R$ .....	16
4.1.1.3 Objetos.....	18
4.1.2 Gráficos BÁSICOS de control estadístico .....	18
4.1.2.1 Diagrama De Tallo Y Hoja .....	18
4.1.2.2 Histograma .....	19
4.1.2.3 Diagrama de caja .....	19
4.1.2.4 Diagrama de Pareto .....	21
4.1.2.5 Diagrama Ishikawa: .....	22
4.1.3 Cartas de control .....	22
4.1.3.1 Ventajas de las cartas de control .....	23

4.1.3.2	Paso a paso para elaborar cartas de control .....	24
4.1.3.3	Cartas de control $\bar{X} - R$ .....	25
4.1.3.4	Cartas de control $\bar{X} - S$ .....	26
4.1.3.5	Carta de control <i>X-One</i> para mediciones individuales .....	26
4.1.3.6	Carta P .....	27
4.1.3.7	Carta np: .....	28
4.1.3.8	Carta C.....	29
4.1.3.9	Carta U.....	30
4.1.4	Análisis de la capacidad y desempeño de los procesos .....	31
4.1.4.1	Índices de capacidad Cp y Cpk .....	32
4.1.4.2	Índices de desempeño Pp y Ppk: .....	34
5	DISEÑO METODOLOGICO.....	36
6	DESARROLLO Y RESULTADOS.....	37
6.1	Fase investigativa .....	37
6.1.1	Identificación de ejes temáticos .....	37
6.1.2	Revisión bibliográfica: .....	37
6.1.3	Panorama básico de interfaces y comandos.....	38
6.2	Fase operativa.....	38
6.2.1	Diseño y estructura de la plantilla guía .....	38
6.2.2	Contextualización de comandos de R con los temas de cada guía.....	41
6.2.3	Planificación de la metodología y actividades de la guía.....	41
6.2.4	Desarrollo del código de la guía en los programas correspondientes .....	42
6.2.5	Desarrollo de las instrucciones para la entrega del informe .....	43
6.3	Fase de verificación y ajustes .....	44

6.3.1	Revisión del borrador inicial .....	44
6.3.2	Desarrollo de pruebas.....	44
6.3.3	Reajustes pertinentes .....	46
7	CONCLUSIONES .....	47
8	RECOMENDACIONES.....	48
	ANEXOS .....	49
	Anexo A. Guías de aprendizaje .....	50
	Anexo B. Videos de pruebas a las guías .....	164
	BIBLIOGRAFÍA.....	165

## LISTA DE FIGURAS

	pág.
Figura 1. Visión esquemática del funcionamiento de R.....	17
Figura 2. Diagrama de tallo y hoja de calificaciones. ....	18
Figura 3. Histograma elaborado en <i>Rstudio</i> .....	19
Figura 4. Diagrama de caja elaborado en <i>Rstudio</i> del dataset <i>Mtcars</i> . ....	20
Figura 5. Representación gráfica del <i>principio de Pareto</i> . ....	21
Figura 6. Diagrama Ishikawa elaborado en <i>Rstudio</i> .....	22
Figura 7. Diagrama de flujo de análisis en elección de carta de control .....	23
Figura 8. La carta $\bar{X}$ detecta cambios significativos en la media del proceso .....	25
Figura 9. La carta R detecta cambios en la amplitud de la dispersión .....	25
Figura 10. Cartas de control $\bar{X}$ y S elaboradas en <i>Rstudio</i> .....	26
Figura 11. Diagrama de flujo correspondiente a elección de la carta <i>X-One</i> . ....	27
Figura 12. Análisis gráfico de capacidad junto con los casos más comunes.....	31
Figura 13. Primera plantilla elaborada para prácticas de laboratorio.....	39
Figura 14. Estructura y diseño de la plantilla guía.....	40
Figura 15. Puesta en escena de la Interfaz de <i>Rstudio</i> .....	42
Figura 16. Puesta en escena de la Interfaz de <i>Overleaf-Sharelatex</i> . ....	43
Figura 17. Métricas de presentación de informe de laboratorio.....	43
Figura 18. Estudiante de diplomado realizando prueba .....	44
Figura 19. Participante de la Universidad Nacional sede Palmira en prueba .....	45
Figura 20. Autor realizando prueba con participantes de Ecuador. ....	45

## LISTA DE ECUACIONES

	pág.
Ecuación 1. Cálculo de cuartiles .....	20
Ecuación 2. Proporción de artículos defectuosos .....	27
Ecuación 3. Carta P estandarizada .....	28
Ecuación 4. Carta C estandarizada .....	30
Ecuación 5. Índice de capacidad potencial Cp .....	32
Ecuación 6: Cálculo de desviación estándar del proceso estimado .....	32
Ecuación 7. Índice de capacidad de especificación inferior Cpl .....	33
Ecuación 8. Índice de capacidad de especificación superior Cpu .....	33
Ecuación 9: Índice de la capacidad real del proceso Cpk .....	33
Ecuación 10. Índice de desempeño potencial Pp .....	34
Ecuación 11: Desviación muestral total S .....	35
Ecuación 12. Índice de desempeño de especificación inferior Ppl .....	35
Ecuación 13. Índice de desempeño de especificación superior Ppu .....	35
Ecuación 14: Índice de desempeño real del proceso Ppk .....	35

## GLOSARIO

**ARGUMENTO:** Propiedad de una función que permite al usuario la modificación y ajuste de esta de acuerdo a sus requerimientos. (Wickham, 2014)

**CALIDAD:** Hace referencia a que un producto sea adecuado para su uso. De esta manera, la calidad consiste en ausencia de deficiencias en aquellas características en pro de la satisfacción del cliente. (Guitierrez, 2010)

**COMANDO:** Se establece como una orden o directriz que debe ejecutar un programa de computador para desarrollar una tarea en concreto (Alvarez, 2019)

**ESTADÍSTICA:** Rama de las matemáticas cuyo objeto de estudio se enfoca en la variabilidad y los acontecimientos aleatorios que la producen mediante el estudio de las leyes de la probabilidad. (Ocaña, 2017)

**FORO:** Sitio en internet donde se elabora un espacio en el cual las personas discuten en torno a un tema en concreto. (vBulletin, s.f)

**FUNCIÓN:** Se define como un bloque o conjunto fundamental de código que ejecuta una operación en concreto. (Wickham, 2014)

**GRÁFICO:** Hace alusión a una forma de representar un conjunto de datos en forma de símbolo tales como barras, líneas o porciones como en un gráfico circular. (Jensen & Anderson, 1992)

**IDE:** Un *Entorno de Desarrollo Integrado* es una aplicación cuya función es servir para la construcción de aplicaciones. (Lozano & Ramos, 2000)

**LATEX:** Es un sistema de composición de documentos el cual se utiliza ampliamente en la elaboración de textos de índole científica. (Lamport, 1986)

**OVERLEAF-SHARELATEX:** Editor de texto *Latex* colaborativo en la nube, utilizado para creación de documentos de índole científica (Leek, 2016)

**QCC:** *Quality Control Charts* es un paquete diseñado para *R* enfocado al análisis y control estadístico. (Scrucca, Snow, & Boomfield, 2017)

**R:** Sistema y lenguaje de programación enfocado para análisis estadístico y gráfico. (Paradis, 2003)

**RSTUDIO:** Interfaz que facilita al usuario la elaboración la programación en lenguaje *R*. Se puede trabajar en las versiones tanto para escritorio como web. (RStudio, s.f)

**SOFTWARE:** Conjunto de instrucciones que dictaminan a un computador la ejecución de respectivas tareas de acuerdo al diseño y propósito con el cual fue elaborado. (Technopedia, 2020)

## RESUMEN

El presente proyecto de grado se desarrolló con la misión de elaborar un compendio de guías de aprendizaje, para gráficos de control haciendo uso del software estadístico *R*. La antesala fundamental es una revisión a los contenidos y temáticas trabajados en la asignatura de control de calidad materias afines y nexos, se realizó una labor investigativa de los comandos, funciones y paquetes con los que contaba *R* para poder llevar este fin a cabo.

Enseguida, se realiza una investigación en pro de desarrollar o encontrar luces de la existencia de una plantilla, en la cual, se lograsen plasmar las pautas de aprendizaje recolectadas. Una vez establecido el formato inicial, se procedió con el desarrollo operativo. De esta manera, se elaboraron y probaron un compendio de guías de aprendizaje, las cuales abordan los principales gráficos de control utilizados en el análisis estadístico de la calidad. Este menester demandó una cantidad considerable de trabajo por parte de investigación y prueba de códigos de manera autónoma, para poder así comprobar la efectividad y confiabilidad de los comandos elaborados. Tan pronto como se elaboraban cada una de las guías se sometían a revisión por parte de la directora del proyecto en pro de identificación de oportunidades de mejora. En cuanto esta etapa finalizaba para todo el compendio de material estas se sometieron a las pruebas correspondientes.

## INTRODUCCIÓN

El lenguaje  $R$ , enfocando para el análisis estadístico de datos en la actualidad se está convirtiendo en una pauta que está creciendo de forma vertiginosa en la demandas de la sociedad contemporánea. De esta manera, las tendencias empresariales dan puntapiés a los antiguos modelos y software para estos menesteres, con lo cual, se está adoptando una nueva doctrina y filosofía de trabajo en la cual los futuros profesionales deberán contar con conocimientos sólidos y flexibles que les permitan adaptarse a las actuales demandas del mercado.

Es allí donde entra en escena la cimentación de unas buenas bases establecidas desde los programas académicos, los cuales permitan al estudiante una rápida, concisa pero efectiva y potente capa de habilidades y conocimientos para que este pueda desenvolverse con fluidez y gracia ante los futuros contratiempos que la sociedad, con sus siempre imperantes y veloces avances en pro del desarrollo jalona la cultura profesional y académica. De ser así, los individuos que logren dominar, comprender, ordenar y canalizar el amplio *maremágnum* de conocimientos que el apoteósico mundo de  $R$  contempla, lograrán ser aquellos que puedan suplir las demandas siempre exigentes de los cambiantes mercados, llevando esto con altos niveles de excelencia y calidad en el análisis de la información.

Para lograr este objetivo, es fundamental que los futuros profesionales cuenten con un conjunto de herramientas que, de manera compacta y armoniosa, dispongan una ruta de aprendizaje, que permita con la mayor efectividad posible la aprehensión de estos saberes de tal manera que logren un apalancamiento en pro de sus intereses y el de todos, obteniendo ventajas competitivas y abriendo puertas a las oportunidades. Este menester, tiene una importancia vital en el sentido de que, se logra ampliar el saber en esta nueva tendencia que respecta al lenguaje  $R$  en el objetivo colectivo de ampliar la verdad y extender el progreso.

El material elaborado trata de la manera más práctica y concisa posible, la explicación de las temáticas respectivas y de los códigos a ejecutar. Se espera que, este compendio logre ser de utilidad para los estudiantes y docentes del programa de ingeniería industrial de la *Fundación Universitaria Los Libertadores*.

## 1 PLANTEAMIENTO DEL PROBLEMA

En tiempos anteriores, en la *Fundación Universitaria Los Libertadores*, las temáticas concernientes al control estadístico de calidad, sólo se basaban en instruir al estudiante en el concepto respectivo, pero no en aplicarlo en una herramienta en concreto. Eso ha permitido que, si bien este lograra tener presente los conceptos, no pudiese apropiarlos de una manera eficaz y tener el dominio de aquellas pautas para poder desarrollar la aplicabilidad correspondiente.

Por otro lado, las actuales tendencias hacia el estudio de *BigData*, requerían de una manera mucho más efectiva para poder tratar una cantidad apoteósica de información, de tal manera que, su manejo y análisis fuese lo más efectivo posible. Con lo cual, se inició una transición y adecuación de los sistemas de cómputo en pro de que estos contasen con los software *R* y *Rstudio*. De esta manera, las asignaturas que tengan que utilizar esta clase de herramientas para la analítica de datos empiezan a hacer su proceso de adaptación de parte de las nuevas exigencias académicas. Sin embargo, la aún más grande cantidad de información sobre el uso de funciones, estructuras y argumentos de los códigos que trabajan en lenguaje *R*, aún se encuentran alejados de la literatura en español, lo cual hace que su acceso sea limitado y más importante aún, la forma de aprender estos conceptos se encuentra sumergida en una vasta entropía. De esta manera, en base a esto se plantea: **¿De qué manera se pueden enseñar en el alma mater a elaborar gráficas de control con el uso del software *R* de una manera clara, concisa y ordenada?**

## 2 JUSTIFICACIÓN

Actualmente, la *Fundación Universitaria Los Libertadores* pretende implementar el lenguaje *R*, como un software que contribuya en las prácticas académicas de los estudiantes que cursen asignaturas tales como control de calidad o que, requieran realizar seguimientos a procesos medibles, para su respectivo análisis y visualización de información mediante técnicas estadísticas. Por este motivo, la intención de este proyecto es diseñar y desarrollar herramientas como guías de aprendizajes, que faciliten el uso adecuado de lenguaje *R* y su respectiva interpretación por parte de los estudiantes y docentes en cada sesión de clase. De esta manera, se logra fortalecer la aplicabilidad de los conceptos vistos durante estas, y al mismo tiempo, se centra en los procesos de aprendizaje de los estudiantes en las áreas que les competen, desarrollando habilidades que se requieren ya en el contexto empresarial a nivel mundial.

El desarrollo de las guías de aprendizaje, proporciona a los estudiantes diferentes rutas que faciliten el uso del lenguaje *R* y su comprensión, optimizando el tiempo establecido en cada sección de clase, además facilita el seguimiento y la evaluación por parte de los docentes.

## 3 OBJETIVOS

### 3.1 OBJETIVO GENERAL.

Diseñar un paquete de guías de aprendizaje, las cuales sirvan como herramientas para el desarrollo de prácticas académicas, relacionadas con control de calidad por medio del software *R*.

### 3.2 OBJETIVOS ESPECÍFICOS.

- Revisar los procedimientos con los que cuenta el software *R* para la generación de gráficos de control.
- Diseñar una plantilla que permita establecer una pauta en el desarrollo de prácticas de laboratorio en el alma mater.
- Validar la efectividad y confiabilidad por parte de las guías de aprendizaje desarrolladas, mediante pruebas con un nicho de participantes en concreto.

## 4 MARCO REFERENCIAL

### 4.1 MARCO TEÓRICO

**4.1.1 LENGUAJE R:** Es un lenguaje de programación basado en el antiguo lenguaje S. La idea fundamental de *R* es servir de utilidad en pro del análisis estadístico de datos con lo cual, abarca un gran *maremágnum* de funciones y paquetes desarrollados netamente para este fin. La ventaja que tiene *R* con respecto a otros lenguajes de programación como *Java*, es que cuenta con una estructura que permite que, la elaboración de los códigos requiera pocas líneas de comando y sea más efectiva debido a que, no se requiere hacer la programación de una función como un cálculo de una desviación estándar, o alguna prueba de normalidad y después poner esta función en escena debido a la amplia variedad de funciones y paquetes, estas ya están integradas en unas pocas línea de comando.

El entorno de trabajo de *R* más conocido como ***Rstudio*** es un software libre bajo los términos de la Free Software Foundation 's con Licencia Pública General de GNU en forma de código fuente. Compila y se ejecuta en una amplia variedad de sistemas operativos abarcando desde los clásicos *Windows* y *MacOS* hasta *FreeBSD* y *Linux*.

**4.1.1.1 EL ENTORNO R:** *R* es un conjunto integrado de instalaciones de software para la manipulación de datos, el cálculo y la visualización gráfica. Incluye:

- Una instalación efectiva de manejo y almacenamiento de datos.
- Un conjunto de operadores para cálculos en matrices.
- Una colección grande, coherente e integrada de herramientas intermedias para el análisis de datos.
- Facilidades gráficas para el análisis de datos y visualización en pantalla
- Un lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas por el usuario y facilidades de entrada y salida.

*R*, como un entorno de programación, se desarrolla mediante librerías o también llamados ***paquetes***. Estos completan el lenguaje con nuevos desarrollos previos para distintas áreas de análisis estadísticos y gráficos. (Santana & Farfán, 2014)

**4.1.1.2 FUNCIONAMIENTO DE R:** Este es un lenguaje cuya escritura es fácil e intuitiva sin embargo, esta requiere de unas pautas fundamentales para que los códigos se ejecuten correctamente debido a que, las funciones deben estar estrictamente escritas dentro de paréntesis y separar cada argumento a añadir mediante el uso de comas.

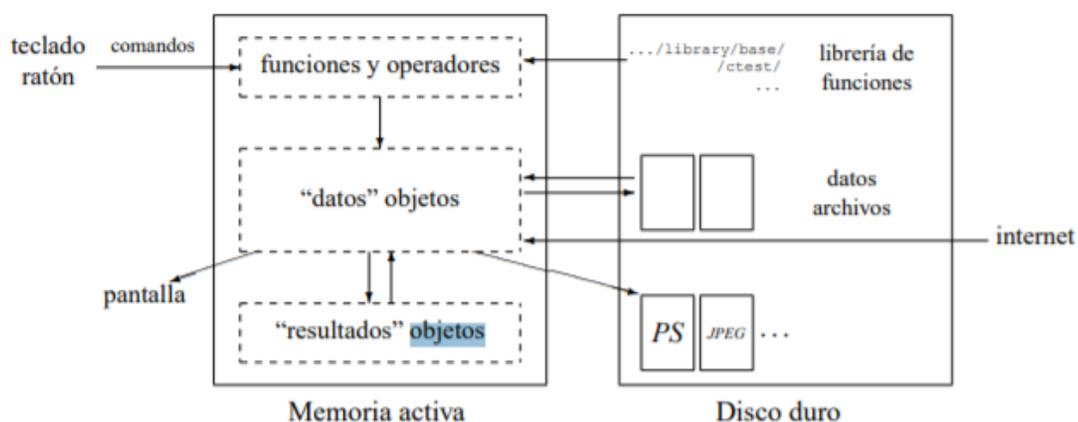
*R* es un lenguaje *Orientado a Objetos*, lo cual significa que las variables, datos, funciones, resultados, y demás, se guardan en la memoria activa del computador en forma de objetos con un nombre puntual. El usuario puede modificar o manipular estos objetos con operadores que pueden ser aritméticos, lógicos, y comparativos y además de otras funciones integradas que también clasifican como objetos.

Por otra parte, *R* utiliza funciones para realizar una acción. Una función es una instrucción diseñada para ejecutar una operación en específico, por ejemplo *mean()* tiene el propósito de calcular el promedio de un conjunto de datos en concreto. Cabe añadir que, para utilizar una función deben especificarse unos argumentos los cuales son los que se escriben dentro de los paréntesis. Desde otro ángulo, existen también algunas configuraciones especiales a ciertos argumentos los cuales deben ingresarse de una forma especial, es aquí donde entran en escena los tipos de variables los cuales son:

- **Lógicos:** se dictaminan como *TRUE* o *FALSE*.
- **Numéricos:** Se establecen vectores cuya estructura inicia con la letra “c” y dentro de los paréntesis se escribe cada número de forma individual separado por una coma. Un ejemplo sería un vector con los valores del 1 al 5 cuya estructura quedaría como *c(1,2,3,4,5)*.
- **Caracteres:** Se dictaminan palabras entre comillas para indicar el nombre a una parte de un gráfico o un conjunto de datos más conocido como ***dataframe***.

**Figura 1.** Visión esquemática del funcionamiento de R

(Paradis, 2003)



**4.1.1.3 OBJETOS:** En términos generales, todos los elementos que maneja *R* son objetos: un valor numérico, un vector, una función, una base de datos y un gráfico. (Venables & Smith, 1990). Algunos aspectos importantes a tener en cuenta de los objetos son:

- El nombre de un objeto debe comenzar con una letra. Ya a partir de adelante puede incluir letras, dígitos (0-9), y puntos (.) (salvo operadores aritméticos o lógicos). *R* discrimina entre letras mayúsculas y minúsculas para el nombre de un objeto, de tal manera que “x” y “X” hacen alusión a objetos diferentes (inclusive bajo Windows).
- El modo de asignación de valores a un objeto es mediante los símbolos de igual “=” y una flecha con guion hacia la derecha “<-”. En este último, es importante tener en cuenta que una vez se ejecuta esta asignación, el objeto queda guardado junto al conjunto de datos o dataframes utilizados en la parte superior derecha de la pantalla. De esta manera *R* logra mantener al usuario los elementos que este crea de una manera ordenada y concisa.

#### 4.1.2 GRÁFICOS BÁSICOS DE CONTROL ESTADÍSTICO

**4.1.2.1 DIAGRAMA DE TALLO Y HOJA:** Es una herramienta gráfica que permite de una forma sencilla visualizar un conjunto de datos. Este gráfico se dio a conocer por el matemático estadounidense **John Tukey** en su libro **Análisis Exploratorio de Datos** del año 1993. La creación de esta clase de diagramas es fácil de realizar. Debe seguir estos pasos:

1. Realizar dos columnas y nombrarlas como “Tallos” y “Hojas” respectivamente.
2. En la primera columna coloque el primer dígito (decena) de izquierda a derecha y en la segunda columna debe colocar las unidades que compartan al dígito de decena que se encuentra en la columna “Tallos”.
3. Puede evidenciar el uso del diagrama de tallo y hoja en la **Figura 2** en la cual se han ordenado las calificaciones obtenidas por 30 estudiantes en un examen.

**Calificaciones:** 63, 58, 61, 52, 59, 65, 69, 75, 70, 54, 57, 63, 76, 81, 64, 68, 59, 40, 65, 74, 80, 44, 47, 53, 70, 81, 68, 49, 57 y 61. (mrbartonmaths, 2005)

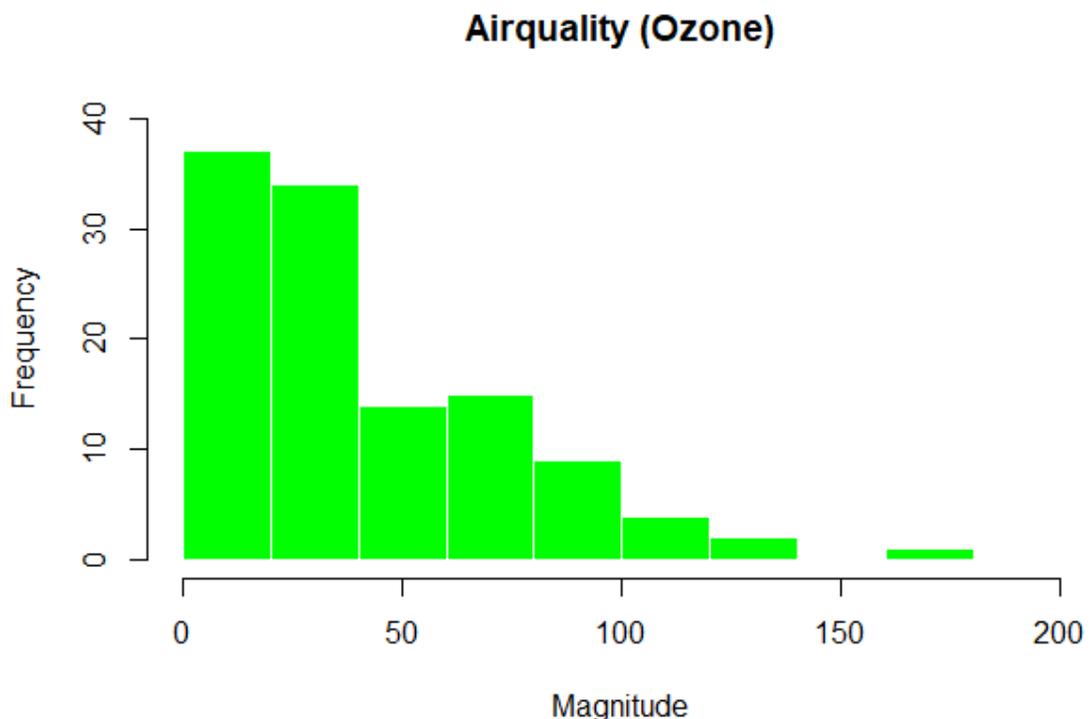
**Figura 2.** Diagrama de tallo y hoja de calificaciones.

Tallos	Hojas
4	0479
5	82947937
6	3159348581
7	50640
8	101

Las interpretaciones del diagrama de tallo y hoja pueden ser diversas dependiendo del contexto de la muestra o población observada. Sin embargo, a rasgos generales se pueden evidenciar dos comportamientos principales: La grafica presenta **simetría**, es decir, si las proporciones del grafico son iguales en ambos lados. Por otro lado se presenta el caso de ser **asimétrica** o **sesgada** con lo cual se observaría una tendencia hacia los datos mayores o menores bien sean en la parte inferior o superior del diagrama. Además de esto, pueden evidenciarse muchas otras medidas estadísticas y novedades en la información con el comportamiento que están presentando los datos tales como valores atípicos, curtosis o apuntamiento de la distribución, entre otros.

**4.1.2.2 HISTOGRAMA:** Es una representación gráfica, que agrupa datos de una forma muy similar a lo que se hace en un diagrama de tallo y hoja, sin embargo, este presenta la información de forma más completa y detallada, con lo cual, permite hacer un análisis mucho más intensivo de lo que se está presentando.

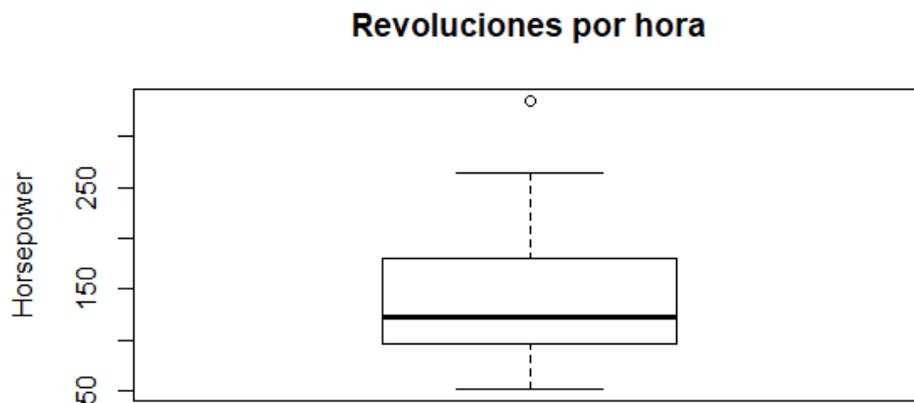
**Figura 3.** Histograma elaborado en *Rstudio*



**4.1.2.3 DIAGRAMA DE CAJA:** Un diagrama de caja, o también conocido como diagrama de caja y bigote, es una representación que permite evidenciar de una forma sencilla la distribución de los datos acudiendo a sus cuartiles. Esta herramienta visual consta de una caja, en la cual la parte inferior pertenece al primer cuartil, la línea central corresponde a la mediana o segundo cuartil, y el borde

superior de la caja representa al tercer cuartil. También se dibujan unos límites para los valores máximos y mínimos de la población; estos datos son los llamados “bigotes”. Hay casos en los que, se pueden apreciar datos más allá de estas demarcaciones, los cuales son valores atípicos de la muestra o la población evaluada. La forma para elaborar diagramas de caja se exponen en los siguientes pasos.

**Figura 4.** Diagrama de caja elaborado en *Rstudio* del dataset *Mtcars*.



1. Ordenar el conjunto de datos de menor a mayor.
2. Obtener los cinco puntos fundamentales para la elaboración del gráfico:
  - Identificar los valores **máximos** y **mínimos** del conjunto de datos.
  - Obtener la **mediana** correspondiente. Tener presente que, si el número valores es impar, esta medida es justamente la mitad de los datos. Por otro lado, si el número de valores es par, será la media de los dos números que se encuentran en la mitad justamente.
  - Obtener los **cuartiles** del conjunto de datos. Si bien no hay un consenso en cuanto a un sólo método para hallar estas medidas de posición, una ecuación efectiva para este menester es:

**Ecuación 1.** Cálculo de cuartiles

$$L = \frac{k}{100}$$

(Triola, 2004)

Donde **L** es el valor del percentil que se desea obtener; **k** es el número de ubicación del percentil (recuerde que los percentiles son medidas de posición que dividen el

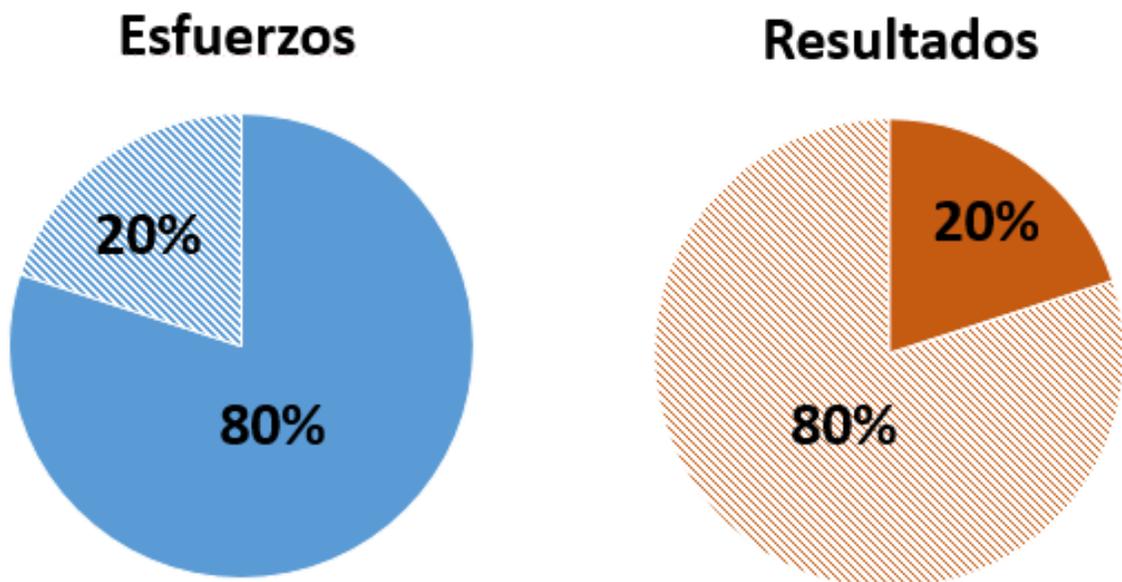
conjunto de datos en 99 puntos. En ese orden de ideas, los cuartiles corresponden a 25, 50 y 75). Finalmente,  $n$  es el número de valores del conjunto de datos.

3. Dibujar la caja correspondiente, basándose en una escala de medida, ubicando los cinco puntos encontrados en el inciso anterior, realizando una caja con los cuartiles y unas extensiones hasta los valores máximos y mínimos. Se aprecia en la **Figura 4** un ejemplo de diagrama de caja:

**4.1.2.4 DIAGRAMA DE PARETO:** Es una herramienta gráfica, que permite realizar un análisis exhaustivo en el que, se dictamina el enfoque hacia los problemas vitales debido a que, es imposible abordar todos los problemas al mismo tiempo. Se basa en el *Principio de Pareto*, creado por el economista *Vilfredo Pareto*, el cual establece que el 80 % de los resultados son producidos por el 20 % de las causas.

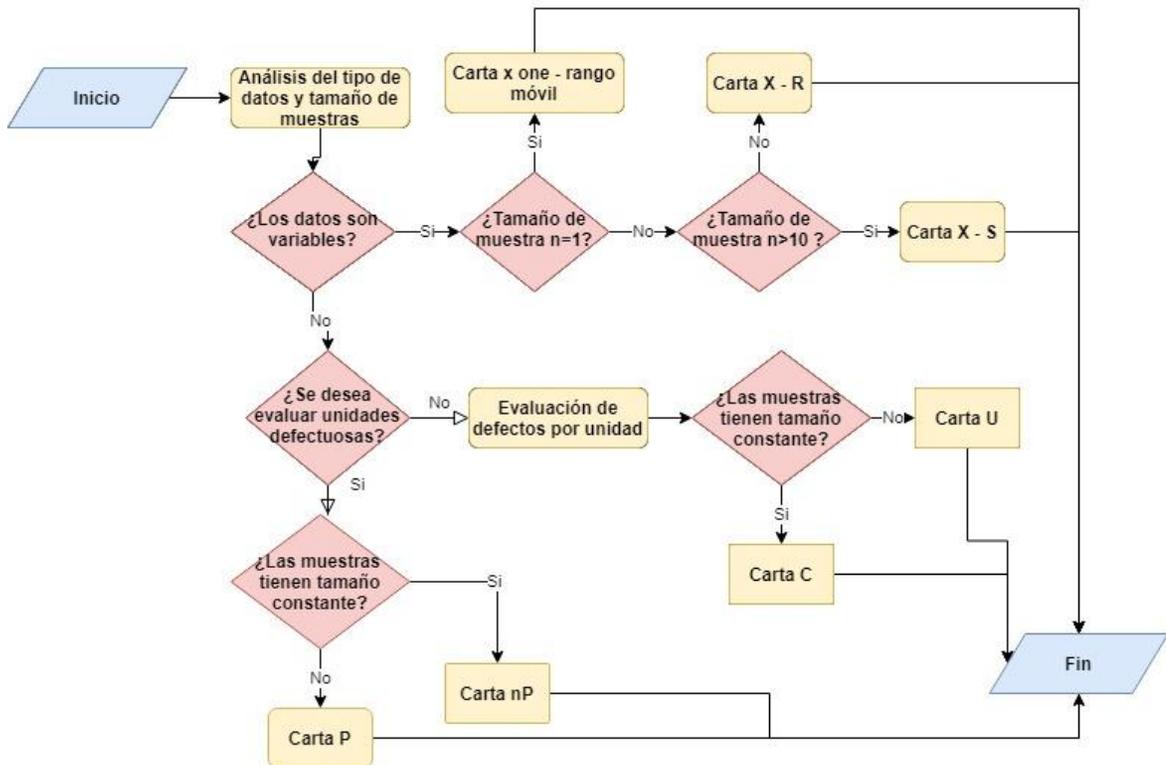
Los diagramas de Pareto resultan ser una poderosa herramienta, en lo que respecta a análisis exhaustivos y eficientes, permitiendo de forma vertiginosa establecer un panorama concreto, en pro de elaborar planes de acción para realizar mejoras en los procesos.

**Figura 5.** Representación gráfica del *principio de Pareto*.





**Figura 7.** Diagrama de flujo de análisis en elección de carta de control



- **Cartas de control para variables:** Permiten la medición de manera cuantitativa, de una magnitud en concreto (temperatura, longitud, viscosidad, entre otros). La idea es, realizar un control al cumplimiento de los requerimientos en base a una métrica especificada.
- **Cartas de control por atributos:** Estas permiten dictaminar si el proceso está bajo control estadístico, en base a la conformidad y no conformidad de este en función de ciertos parámetros cualitativos.

El concepto de la carta de control fue desarrollado en 1920 por el físico, ingeniero y estadístico estadounidense **Walter A. Shewhart** de *Bell Telephone Laboratories*. (Montgomery, 2004)

#### 4.1.3.1 VENTAJAS DE LAS CARTAS DE CONTROL

1. Son una herramienta magistral en pro de la productividad ya que dimite en el hecho de reducir los desperdicios.
2. Previenen defectos.

4. Evitan el ajuste innecesario del proceso.
5. Son herramientas precisas debido a que, identifican con claridad las causas fortuitas y asignables. Esto es fundamental ya que, un operario no tiene la sensibilidad para reconocer esta clase de ruido blanco.
6. Brinda información relacionada con la capacidad del proceso.

**4.1.3.2 PASO A PASO PARA ELABORAR CARTAS DE CONTROL:** El monitoreo del control estadístico de calidad, a través de las cartas de control, debe seguir los siguientes pasos:

1. Calcular los límites de control y *a posteriori* realizar el gráfico correspondiente.
2. Realizar un análisis a los puntos de las cartas de control, verificando las causas fortuitas y revisando detalladamente las causas asignables del proceso. Si se encuentran las causas asignables que están ocasionando inconvenientes en el proceso estas deben ser eliminadas.
3. Una vez se han eliminado las causas asignables, se deben calcular nuevamente los límites de control y enseguida, elaborar las respectivas cartas. En *Rstudio* para poder realizar este paso se deberá realizar una modificación al conjunto de datos con el que se ha trabajado, esto es, elaborar un nuevo dataframe, con los datos extraídos. Esto se puede realizar haciendo uso de la línea de comando de la forma:

$$b <- a[-c(n1, n2), ]$$

Donde ***b*** es el nombre asignado al dataframe con los datos extraídos y ***a*** es el conjunto de datos originales. Para extraer los datos, se debe escribir un vector que contenga la posición de los datos que se desean eliminar entre llaves precedido del signo menos (-) y separado de una coma. Esta estructura dentro de las llaves, lleva la forma [filas, columnas] con lo que, se está indicando eliminar esas filas en cuestión, pero ninguna columna. Es por esta razón que, en la estructura evidenciada anteriormente no se coloca nada después de la coma.

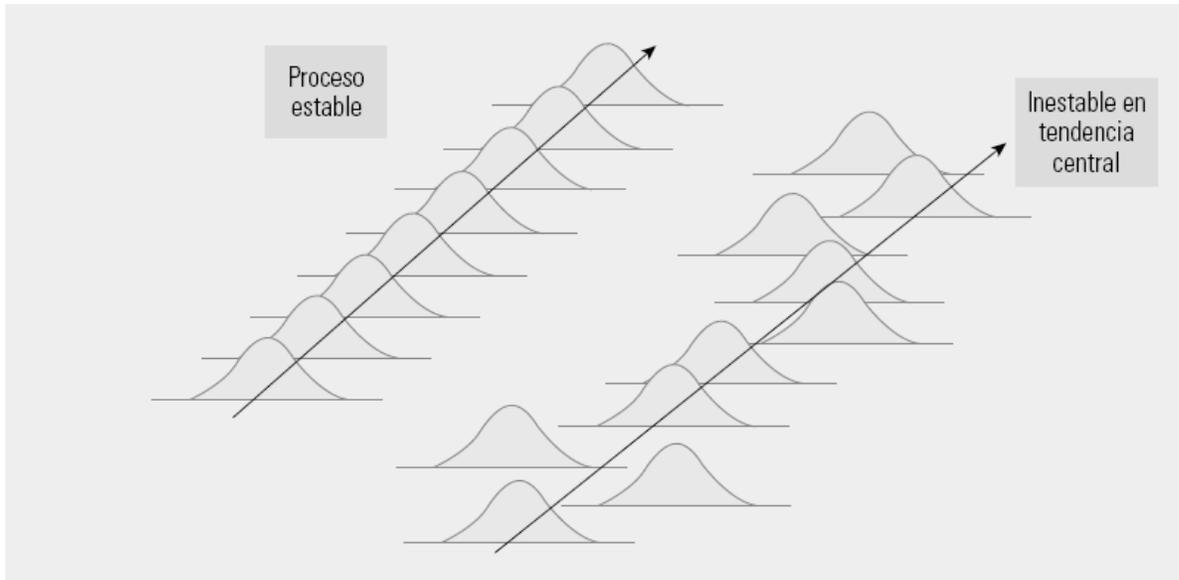
Ahora bien, si se desea eliminar un segmento de datos puntuales, se ha de requerir hacer uso de dos puntos (:) en donde se debe colocar el dígito de las posiciones de inicio y fin del intervalo  $n_1$  y  $n_1$  respectivamente. De esta forma, el comando a usar queda con el siguiente aspecto:

$$b <- a[-c(n1:n2), ]$$

Una vez realizado este procedimiento, se procede a repetir el ploteo de los gráficos con el nuevo dataframe.

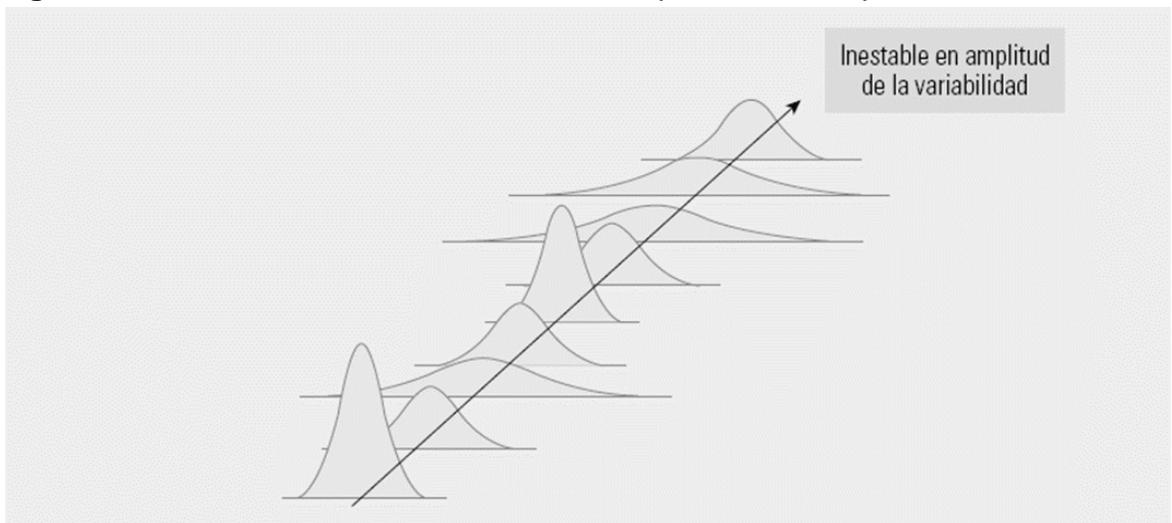
#### 4.1.3.3 CARTAS DE CONTROL $\bar{X} - R$

**Figura 8.** La carta  $\bar{X}$  detecta cambios significativos en la media del proceso



(Gutierrez, 2010)

**Figura 9.** La carta R detecta cambios en la amplitud de la dispersión



(Gutierrez, 2010)

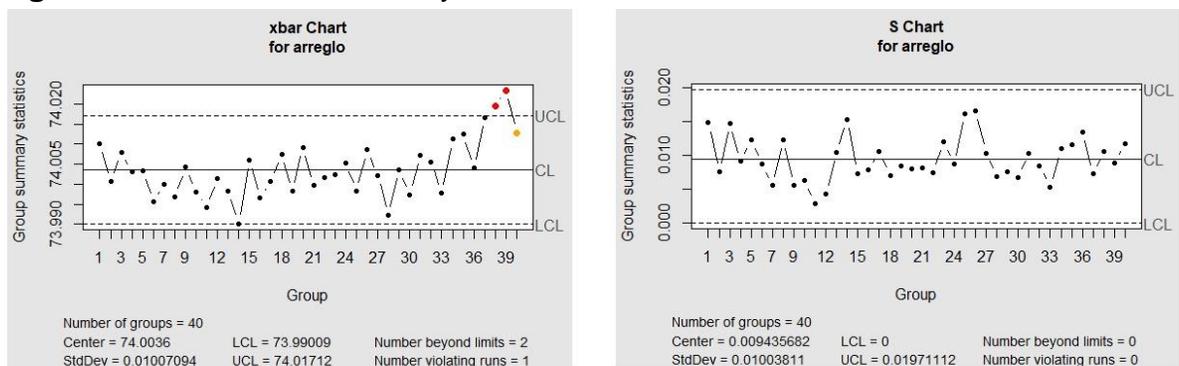
Las cartas de control para medias  $\bar{X}$  y rangos R trabajan de forma armoniosa, con el fin de monitorear procesos en los que se utilicen tamaños de muestras  $n$  menores a 10. La carta  $\bar{X}$  permite el monitoreo de la media de las muestras en el proceso

mientras que, la carta R permite observar la variabilidad de este. Es importante realizar la elaboración de estos gráficos de control en conjunto, debido a que existen procesos en los que, la media de las muestras se encuentra bajo control sin embargo su variabilidad es inestable, y otros en los que, la variabilidad está controlada y el promedio de las muestras puede estar siendo afectado. En la apreciación de la **Figura 8** y también en **Figura 9** esto se puede apreciar más claramente

**4.1.3.4 CARTAS DE CONTROL  $\bar{X}$ -S:** Las cartas de control  $\bar{X}$  y S, son el otro caso práctico de uso para monitoreo de procesos mediante variables. Estas cartas de control se utilizan en los siguientes casos:

1. El tamaño de las muestras  $n$  es superior a 10 o 12.
2. Se presenta una variabilidad en el tamaño de las muestras, es decir, no todas las muestras presentan el mismo tamaño, lo cual hace que, los límites de control también cambien cada que se grafica un punto en las cartas de control.

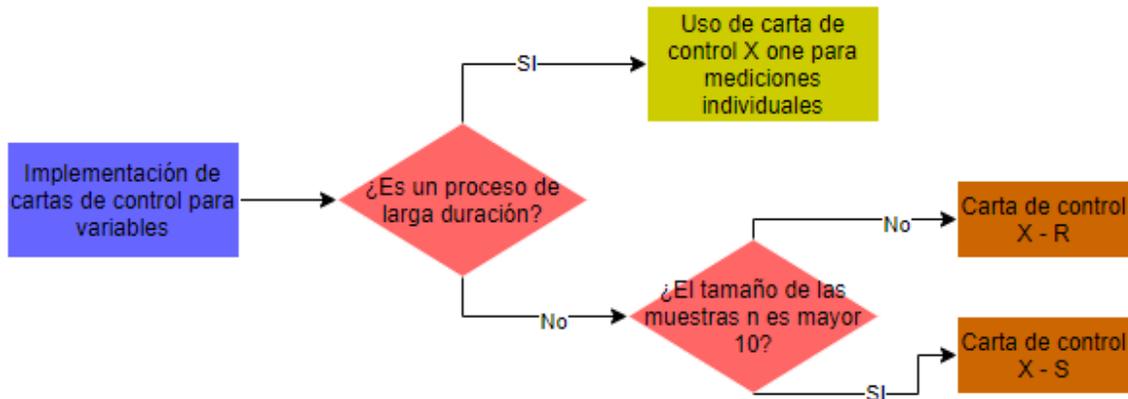
**Figura 10.** Cartas de control  $\bar{X}$  y S elaboradas en *Rstudio*



**4.1.3.5 CARTA DE CONTROL X-ONE PARA MEDICIONES INDIVIDUALES:** La carta de control X-One para mediciones individuales, hace parte del grupo de cartas de control para variables. Se basa en, como su nombre lo indica, hacer el respectivo análisis a cada uno de los productos resultantes. Esto se toma en cuenta, cuando los procesos son demasiado lentos, en los cuales se puede tardar un día o más para tener una muestra. Algunos casos pueden ser:

- Productos químicos.
- Industria de licores.
- Variables administrativas como índices de productividad, consumo de servicios públicos.

**Figura 11.** Diagrama de flujo correspondiente a elección de la carta X-One.



Por otra parte, a pesar de que la carta de mediciones individuales muestra un panorama relevante del proceso, existe una carta que complementa su función: **la carta de rango móvil**. Este gráfico funciona de una manera similar a la carta de rangos. La diferencia radica en los estadísticos  $D_3$  y  $D_4$  que tendrán tamaño 2 (Esto a que cada dato del rango móvil se obtiene de su consecutivo (Gutierrez, 2010)).

**4.1.3.6 CARTA P:** Las cartas de control por atributos, son la otra clasificación para el análisis estadístico por medio de cartas de control. Estas se desarrollan, con el fin de evaluar procesos en los que no es posible cuantificar mediante una magnitud si no mediante parámetros tales como *conforme* o *no conforme* (por ejemplo, al cuantificar la magnitud de la rayadura de una pieza, o el número de defectos que esta cuenta). La carta para proporción de artículos defectuosos o más conocida como **Carta P** permite establecer una proporción  $p_i$  de artículos defectuosos  $d_i$  en un tamaño determinado de muestra o lote  $n_i$  de tal manera que la proporción se obtiene a partir de:

**Ecuación 2.** Proporción de artículos defectuosos

$$p_i = \frac{d_i}{n_i}$$

(Gutierrez, 2010)

- **Cálculo de los límites de control:** A diferencia de las cartas de control para variables, los límites de control en estos gráficos no dependen de una constante, sino más bien del tamaño de la muestra  $n_i$  para cada proporción. Esta carta de control sigue una distribución binomial.

- **Variaciones de la carta P:** Existen casos en los que el tamaño de la muestra  $n_i$  es variable (es decir, no todas las muestras tendrán el mismo tamaño). Con lo cual se plantean dos alternativas:

1. Utilizar el tamaño promedio de las muestras  $\bar{n}$ . Esto es recomendable cuando el número de muestras no sobrepase el 20 % del total del tamaño de estas.

2. Utilizar límites de control variables, es decir, calcular para cada tamaño de muestra  $n_i$  los límites de control respectivos, sin embargo, hay algunos casos en los que esto puede resultar inconveniente a la hora de evaluar la carta de control. Una alternativa extra para este caso es utilizar la carta  $p$  estandarizada, la cual utiliza como UCL y LCL los valores de 3 y -3 respectivamente. La proporción promedio  $\bar{p}$  se corresponde con los anteriores y cada proporción  $p_i$  se somete a la ecuación:

**Ecuación 3.** Carta P estandarizada

$$Z_i = \frac{p_i - \bar{p}}{\sqrt{\frac{q(1 - \bar{p})}{n}}}$$

(Gutierrez, 2010)

**4.1.3.7 CARTA NP:** La carta de control para el número de artículos defectuosos, es una variación de la carta  $p$ , en la cual, no se grafican proporciones sino el número de defectos. Esto es útil en casos en los que el tamaño de la muestra no sea constante. Al igual que en la carta  $p$ , este gráfico se basa en una distribución binomial.

Estas cartas de control, permiten visualizar la realidad del proceso en base de la forma de muestreo de los datos, con lo que es importante el hecho de la respectiva identificación de causas asignables que estén presentes en el proceso. Otro factor importante a tener en cuenta es el caso en el que se presente un valor negativo en el límite de control inferior con lo cual, habría que revisar con detalle errores en la elaboración de estos parámetros o en su defecto, optar por establecer que este límite tenga un valor de cero ya que, es lógicamente imposible que existan unidades defectuosas negativas.

- **Ventajas, desventajas y usos:** La carta  $np$  resulta de fácil lectura para aquellas personas con un conocimiento básico de estadística ya que, simplemente se están trabajando con valores enteros. Por otra parte, al ser de tamaño de muestra constante, la carta resulta ser prácticamente la misma a la carta  $p$ , con lo que, para convertir una carta de proporción en una de unidades defectuosas, basta dividir la

escala de la carta en el tamaño de las muestras. Por lo tanto, se puede decir que el criterio para escoger una carta de otra, depende de la disposición del análisis deseado entre proporción o unidades. Esto partiendo de la base que se menciona anteriormente de que se trate de tamaños de muestra constantes.

Por otra parte, este gráfico presenta la desventaja de que, si bien se puede trabajar esta carta con tamaños de muestras variables, su lectura presenta ser en demasía una inexorable complejidad (Montgomery, 2004). Más aún al requerir un tipo de carta estandarizada, por lo cual, en estos casos, se recomienda mucho más el uso de la carta  $p$ .

**4.1.3.8 CARTA C:** Un artículo no conforme, es una unidad que cumple con una o varias especificaciones dadas para el producto. Cada punto en el que no se cumplen las especificaciones se denomina como defecto o disconformidad sin embargo, dependiendo de la magnitud y naturaleza de este, se puede tomar la decisión de no clasificarlo como *no conforme*, debido a que pueden ser pequeñas imperfecciones que no implican una ineficiencia en el uso del producto. Caso contrario sucede con aquellas unidades que poseen defectos considerables que acarrearán dificultades para el cliente. (Montgomery, 2004)

Existen ocasiones en las que, es mucho más conveniente trabajar con el número de defectos por unidad que simplemente con la proporción no conforme con lo cual, a la hora de medir esta cantidad de defectos en los productos, entra en escena el uso de la carta de control para número de defectos por subgrupo o más conocida como **Carta C**.

La carta de control  $C$  tiene como objetivo analizar el número de defectos por subgrupo, cuando el número del subgrupo se mantiene constante (Gutierrez, 2010) y parte de la base de que los datos se disponen como una distribución de *Poisson*. Por otra parte en cuanto respecta a los casos de uso, esta carta de control aplica en puntos de inspección en los cuales se busca localizar uno o más tipos de defectos relativamente menores o tolerables, de tal forma que, aunque se encuentren estos errores, no se clasifique la unidad como no conforme si no sobrepasa el umbral de defectos establecidos por la organización. Algunos ejemplos de su uso aplican para:

- Número de quejas
- Número de errores
- Cantidad de paros o detenciones.

Se recomienda que, de ser posible se elijan tamaños de subgrupo de forma tal que el número promedio de defectos por subgrupo (línea central) sea mayor que nueve. Por otra parte, si en cada muestra se esperan cero o muy pocos defectos, mucho

menos que nueve, usualmente la carta no es eficaz. En esos casos, se aconseja incrementar el tamaño del *sample* o buscar otras alternativas.

**4.1.3.9 CARTA U:** La carta de control para medición de defectos promedio por unidad, o más conocida como carta *U* se utiliza cuando el tamaño de la muestra o subgrupo no son constantes. Al igual que su antecesora para número de defectos por subgrupo, esta también se ciñe a una distribución de Poisson.

En lo que respecta a los usos y aplicaciones, se aplica igual que la carta *C*, pero aquí se prefiere analizar el número promedio de defectos por artículo o unidad, en lugar del número de defectos por subgrupo. Algunos ejemplos son:

- Número de no conformidades por lote o subgrupo de materia prima recibida donde el tamaño de este varía.
- Cantidad de nuevas infecciones en un hospital por día.
- Cantidad de accidentes por camión de reparto por día.

Al igual que ocurre con su gráfico análogo predecesor, si el tamaño de la muestra *n* es mucho menor que el número recomendado (9), la carta *U* no suele ser útil. En esos casos, hay que buscar incrementar *n* o utilizar otra carta de control. Cabe recalcar que, esta carta es más efectiva cuando se toma para analizar tamaños de subgrupos variables, sin embargo, también es posible utilizarla con muestras constantes calculando el número promedio de subgrupo, lo cual es eficaz cuando en el total de estos se presenta una variación menor al 20 % aproximadamente.

Un aspecto importante es que, esta carta de control presenta de manera análoga los mismos inconvenientes con su hermana la **Carta P**, en lo que respecta cuando el tamaño de los subgrupos no es constante. Para ese caso, también se opta por la opción ya mencionada de una proporción promedio o bien, realizar una carta *U* estandarizada, la cual corresponde en este caso a la **Ecuación 4**.

**Ecuación 4.** Carta C estandarizada

$$Z_i = \frac{u_i - \bar{u}}{\sqrt{\frac{\bar{u}}{n_i}}}$$

(Gutierrez, 2010)

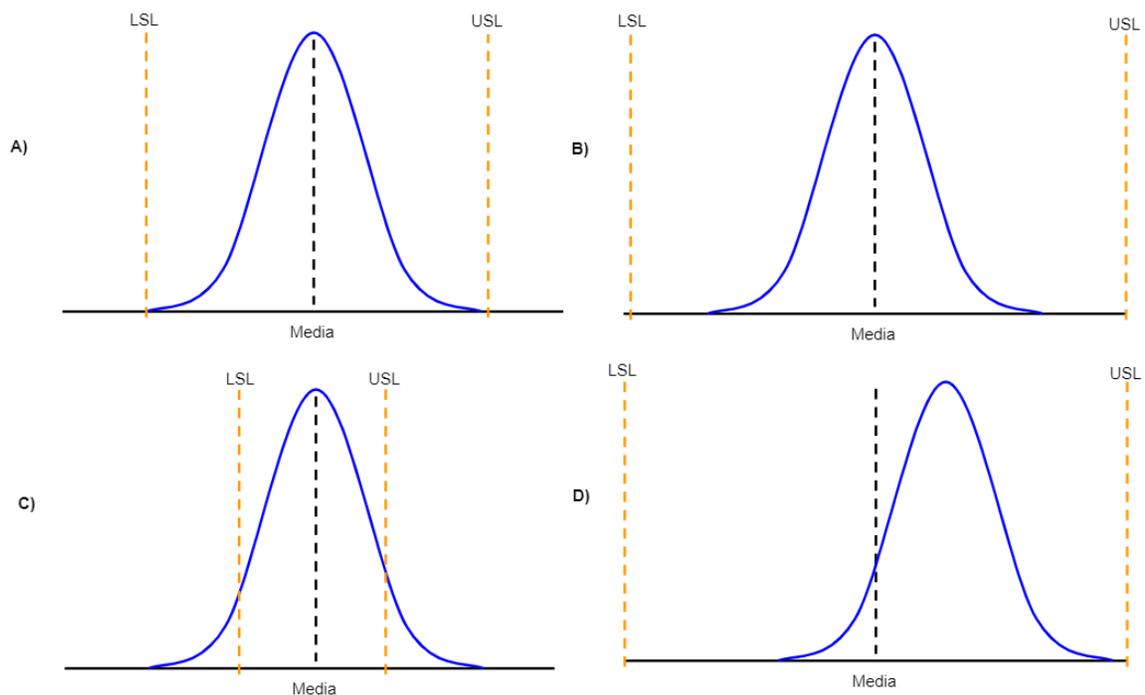
#### 4.1.4 ANÁLISIS DE LA CAPACIDAD Y DESEMPEÑO DE LOS PROCESOS:

Existen casos en donde, las cartas de control no son suficientes para poder realizar un análisis completo del control estadístico de la calidad. Si bien estas pueden establecer un panorama extenso del proceso, es requerido evaluar el comportamiento del producto final. Esto es importante, debido a que puede un proceso estar bajo control estadístico, y sin embargo, no cumplir con los requerimientos demandados por el cliente.

Esta clase de problemáticas se abordan mediante el análisis de la capacidad del proceso, lo cual mide la variabilidad de este de acuerdo a unas métricas establecidas. Estas métricas se evalúan a partir de un parámetro nominal  $N$  al que deben apuntar todos los productos elaborados y unos límites de desviación establecidos que se denominan **límites de especificación**, y corresponden respectivamente al límite de especificación inferior y superior **LSL** y **USL** por sus siglas en inglés.

Todos estos conceptos se pueden apreciar en una distribución normal. En la **Figura 12** se evidencian las situaciones más comunes que se presentan al momento de analizar la capacidad de un proceso:

**Figura 12.** Análisis gráfico de capacidad junto con los casos más comunes.



- El inciso **A)** muestra el caso ideal en el cual, las unidades producidas se encuentran alineadas con el valor nominal que corresponde siempre a la media y además, dentro de los límites de especificación.

- Por otra parte en el inciso **B)** se evidencia límites de especificación demasiado elevados. Es menester, en este caso, ser más exigentes con el análisis y reducir el rango de estos límites.
- El ítem **C)** muestra el caso en el que, si bien el proceso está centrado, este sobrepasa los límites de especificación establecidos. Esto es una señal de alarma, ya que se están produciendo unidades no conformes a los requerimientos del cliente que acarrearán pérdidas considerables a la operación.
- Finalmente se aprecia el inciso **D)** en el cual, la media de las unidades elaboradas no cumple con lo especificado, es decir, el proceso no está centrado.

**4.1.4.1 ÍNDICES DE CAPACIDAD CP Y CPK:** Una forma más precisa de determinar la capacidad del proceso, es mediante un análisis cuantitativo mediante los índices de capacidad, los cuales indican a través de un coeficiente, qué tan bien está el *performance* del proceso.

El primer índice que se establece, es el de capacidad potencial **Cp**, el cual mide el rango o la extensión de las especificaciones en base a una dispersión seis sigma:

**Ecuación 5.** Índice de capacidad potencial  $C_p$

$$C_p = \frac{USL - LSL}{6\sigma}$$

(Gutierrez, 2010)

Donde  $\sigma$  corresponde al cociente entre el rango promedio  $R$  de una muestra de datos y  $d_2$  una constante correspondiente al tamaño de la muestra:

**Ecuación 6:** Cálculo de desviación estándar del proceso estimado

$$\sigma = \frac{R}{d_2}$$

(Gutierrez, 2010)

Si bien es importante medir el proceso mediante el índice **Cp**, debido a que indica si el proceso se encuentra dentro de las especificaciones, este tiene el inconveniente de dejar de lado el valor medio  $\mu$ , lo cual no da la certeza del grado de centralidad del proceso. Para lograr este fin, se calcula el **índice de la capacidad real del proceso Cpk**, el cual si tiene este parámetro en cuenta. Para calcularlo, se deben acudir a los índices de capacidad para la especificación inferior y superior

respectivamente **C<sub>pu</sub>** y **C<sub>pl</sub>**, los cuales calculan la distancia entre la media del proceso con respecto a algún extremo de las especificaciones. Como se están calculando para los extremos inferior y superior respectivamente, sólo se utiliza una variación de 3 sigma. Estos cálculos se evidencian en la **Ecuación 7** y la **Ecuación 8**.

**Ecuación 7.** Índice de capacidad de especificación inferior **C<sub>pl</sub>**

$$C_{pl} = \frac{USL - \mu}{3\sigma}$$

(Gutierrez, 2010)

**Ecuación 8.** Índice de capacidad de especificación superior **C<sub>pu</sub>**

$$C_{pu} = \frac{USL - \mu}{3\sigma}$$

(Gutierrez, 2010)

Una vez hallados estos dos parámetros, el índice **C<sub>pk</sub>** se establece entre el mínimo de los indicadores anteriores.

**Ecuación 9:** Índice de la capacidad real del proceso **C<sub>pk</sub>**

$$C_{pk} = \text{Min}[C_{pu}, C_{pl}]$$

(Gutierrez, 2010)

Este índice refleja un panorama más preciso del proceso, con lo que se deben trabajar de forma conjunta para un efectivo análisis.

En la **Tabla 1**, se especifica la clasificación que tiene cada valor de **C<sub>p</sub>**, con su respectivo dictamen. Estos datos también se utilizan para el índice **C<sub>pk</sub>**:

**Tabla 1.** Valores del índice **C<sub>p</sub>** – **C<sub>pk</sub>** y su indicación.

Valor índice <b>C<sub>p</sub></b> - <b>C<sub>pk</sub></b>	Clase o categoría del proceso	Dictamen (Si el proceso está centrado)
Mayor o igual a 2	Clase mundial	Calidad Seis Sigma
Mayor a 1.33	1	Adecuado
Mayor a 1.44	2	Parcialmente adecuado

**Tabla 2(Continuación).** Valores del índice Cp – Cpk y su indicación

Mayor a 1.44	2	Parcialmente adecuado
Entre 0.67 y menor o igual a 1	3	No adecuado para el trabajo. Requiere serias modificaciones.
Menor a 0.67	4	No adecuado para el trabajo. Requiere con necesidad imperiosa modificaciones
Fuente (Gutierrez, 2010)		

La antesala fundamental para el análisis de la capacidad de los procesos se debe establecer bajo los siguientes criterios:

1. Los datos tienen una distribución normal.
2. El proceso está bajo control estadístico.

Con lo cual, lo primero es establecer la normalidad de los datos ya que, si se calculan los índices de capacidad sin que este inciso se cumpla, se pueden cometer errores de magnitudes graves, en lo que respecta a la incorrecta estimación de la variabilidad. En cuanto al segundo inciso, es aquí donde cobra un papel importante el trabajo de la mano que se realice entre las cartas de control (las de variables son las más utilizadas, sin embargo las cartas de control por atributos también pueden ser un complemento adicional). En el caso de que el proceso no se encuentre bajo control estadístico, se deben calcular los Índices de desempeño del proceso **Pp** y **Ppk**.

**4.1.4.2 ÍNDICES DE DESEMPEÑO PPY PPK:** El Grupo de Acción de la Industria Automotriz (AIAG, por sus siglas en inglés), fue formado en el año de 1991. Este grupo está integrado por los representantes de *Ford*, *General Motors*, *Daimler/Chrysler* y de la *American Society for Quality Control* y uno de sus objetivos era estandarizar los requerimientos para los reportes de los proveedores y, en general de la industria (Montgomery, 2004). El AIAG recomienda usar los índices de la capacidad del proceso  $C_p$  y  $C_{pk}$  cuando el proceso está bajo control estadístico, con la desviación estándar del proceso estimado  $\sigma$ . Cuando el proceso no está bajo control, el AIAG sugiere usar los índices de del desempeño **Pp** y **Ppk**. El índice de capacidad potencial se evidencia en **Ecuación 10** en donde **S** es la desviación muestral del total de los datos la cual corresponde a la **Ecuación 11**.

**Ecuación 10.** Índice de desempeño potencial  $P_p$

$$p_p = \frac{USL - LSL}{6S}$$

(Montgomery, 2004)

**Ecuación 11:** Desviación muestral total  $S$

$$S = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}}$$

(Montgomery, 2004)

De tal manera, las ecuaciones **Ecuación 7**, **Ecuación 8** y **Ecuación 9** se transforman en sus análogas **Ecuación 12**, **Ecuación 13** y **Ecuación 14** que se muestran a continuación:

**Ecuación 12.** Índice de desempeño de especificación inferior  $P_{pl}$

$$P_{pl} = \frac{USL - \mu}{3S}$$

(Montgomery, 2004)

**Ecuación 13.** Índice de desempeño de especificación superior  $P_{pu}$

$$P_{pu} = \frac{USL - \mu}{3S}$$

(Montgomery, 2004)

De la misma manera, el índice  $P_{pk}$  se establece entre el mínimo de los indicadores anteriores.

**Ecuación 14:** Índice de desempeño real del proceso  $P_{pk}$

$$P_{pk} = \text{Min}[P_{pu}, P_{pl}]$$

(Montgomery, 2004)

Muchos autores están en contra del uso de los índices de desempeño, debido a que, no es recomendable canalizar esfuerzos en optimizar un proceso que está alejado de su performance óptimo, sin embargo, resultan convenientes para poder visualizar los resultados respectivos en términos de corto y largo plazo. Esto toma relevancia teniendo presente el hecho que, por el lado de la **Ecuación 6** se calculan estos índices a partir de la desviación de una muestra mientras en la **Ecuación 11** se calcula toda la desviación de los datos.

## 5 DISEÑO METODOLOGICO

Para el satisfactorio desarrollo de este proyecto de grado, se pretende llevar a cabo la siguiente metodología dividida en las siguientes fases:

**Fase investigativa:** Se realizará una labor de estudio teórica e investigativa sobre los diferentes tópicos en cuanto al control estadístico de calidad se refiere y los software *Latex* y *Rstudio*. Esto con el fin de lograr un entendimiento profundo y certero sobre las temáticas con el objetivo de poder plasmar en los documentos los conceptos y metodologías, de forma entendible y agradable para el uso posterior de los estudiantes y docentes. En esta fase se encuentran las actividades:

- Identificación de ejes temáticos.
- Revisión bibliográfica sobre el control estadístico de calidad.
- Identificación y contextualización básico de interfaces y comandos

**Fase operativa:** Una vez realizado el respectivo estudio, se prosigue con la elaboración del código requerido de las guías de aprendizaje en los software respectivos, los cuales se complementan para este fin de forma armoniosa, ofreciendo poderosas herramientas para este menester. Aquí se realizan las siguientes tareas:

- Estructura y diseño de la plantilla guía en *Latex*.
- Contextualización de comandos del software *R* con los temas de cada guía.
- Planificación de la metodología y actividades de cada guía.
- Desarrollo del código de la guía en los programas correspondientes.
- Desarrollo de las instrucciones para la entrega del informe.

**Fase de verificación y ajustes:** Una vez se haya finalizado con el diseño de todas las guías de aprendizaje, estas se verán sometidas a pruebas con grupos de estudiantes para verificar la efectividad de estas, las cuales *a posteriori* serán objeto de ajustes de las mejoras identificadas. Esta fase la comprenden las actividades:

- Revisión del director del proyecto
- Desarrollo de pruebas
- Reajustes pertinentes

## 6 DESARROLLO Y RESULTADOS

### 6.1 FASE INVESTIGATIVA

**6.1.1 IDENTIFICACIÓN DE EJES TEMÁTICOS:** Lo principal fue la identificación y dictamen de los temas que se planteaban desarrollar, mediante las guías de aprendizaje. Inicialmente, se pretendía la idea de un total de 21 guías que abarcarían desde los temas más básicos de elaboración de gráficos de control hasta el análisis multivariable más complejo; toda esta empresa repartida entre dos estudiantes. De esta manera, este abanico de temáticas facilitaría de una forma integral el aprendizaje de estos conocimientos, tanto a los estudiantes de pregrado como de diplomado y posgrado.

De acuerdo a consenso de los docentes, se dictaminó que cada estudiante realizara un set de guías para presentar de forma independiente, lo cual no repercutió en la distribución del menester, sin embargo, al final la otra estudiante por razones personales tomó la decisión de no continuar con esta empresa, con lo que en consenso con la directora del proyecto, se reajustó el compendio de ejes temáticos para que **se desarrollarán un total de 13 guías de aprendizaje** las cuales se ajustaran al *Syllabus* de la asignatura de *Control De Calidad*.

Esto marcó una pauta importante, debió a que, trazó la ruta más clara posible para que este material sirviese como una poderosa herramienta no sólo al estudiante que está adquiriendo las habilidades y pericias correspondientes al uso de gráficos de control y sus diferentes contextos a utilizar, si no también al docente ya que le permite de una manera mucho más práctica y concisa el desarrollo de sus clases.

**6.1.2 REVISIÓN BIBLIOGRÁFICA:** La disposición de un panorama claro permitió en este momento, iniciar con una labor primordial de hacer una de búsqueda informativa en la literatura existente sobre los temas concernientes a control estadístico de calidad, estadística descriptiva y manejo de lenguaje *R*. En lo que respecta al tópico estadístico, esta información fue obtenida realizando búsquedas en bibliotecas tanto públicas como en la del alma mater y también investigando en la red. Esto tuvo un impacto importante ya que las exigencias del proyecto abrieron la puerta a la necesidad palpable de aumentar y profundizar en conocimientos de estadística y calidad, especialmente abordando con mayor impacto cada uno de los conceptos.

El aspecto que tuvo mayor repercusión, fue la búsqueda de información respectiva con *R* y *Rstudio* ya que, debido al auge relativamente reciente de este lenguaje de programación estadístico la *data* en español no se encuentra aún del todo tan desarrollada, como se evidencia en la literatura de lengua inglesa en la cual, se maneja sobre unos contextos mucho más elaborados y prácticos el flujo de información como foros, paginas tipo wiki y un amplio *maremágnum* de manuales

relacionados al uso respectivo de cada una de la librerías con las que cuenta *R* para sus menesteres, abarcando desde la analítica de datos, hasta el diseño de gráficos en movimiento.

**6.1.3 PANORAMA BÁSICO DE INTERFACES Y COMANDOS:** Fue una necesidad imperiosa, tener conocimientos y habilidades que permitiesen la correcta elaboración de las guías y un buen *performance* de programación en estos lenguajes, lo cual exigió un estudio de sus funcionamientos y aplicaciones respectivas. En lo que respecta al estudio y manejo por parte del lenguaje *Latex*, no se presentaron muchos inconvenientes de gran índole, ya que su uso se ha ido desarrollando en el transcurso del pregrado, tanto por exigencias académicas como bien por trabajo autónomo. De esta manera, en lo que se canalizaron esfuerzos fue en el ya creciente uso del lenguaje *R* y sus interfaz, con lo cual, se revisó de forma preliminar los aspectos más fundamentales sobre su instalación, conocimiento de la interfaz de manejo, funciones básicas, manejo de dataframes integrados en el sistema, uso de paquetes y correcta consulta de sus manuales. De esta manera se cimentan las bases para el desarrollo posterior del material respectivo.

## 6.2 FASE OPERATIVA

**6.2.1 DISEÑO Y ESTRUCTURA DE LA PLANTILLA GUÍA:** . Hasta el año 2016, en la facultad de Ingeniería y Ciencias Básicas no se disponía de una plantilla oficial para la elaboración de prácticas de laboratorio y solamente se contaba con los manuales que se encontraban en la biblioteca y de los formatos desarrollados de manera autónoma por algunos docentes de la facultad. Fue hasta el desarrollo del proyecto de grado *Guías de laboratorio y programación del sistema CIM MPS 500 para su uso en prácticas de métodos y tiempos en la Fundación Universitaria Los Libertadores* de las autoras *Karen Andrea Arce Alarcón* y *Laura Natalia Morales Trujillo* en los que se desarrolló una primera versión preliminar de la plantilla que puede ser visualizada en la **Figura 13**. Cabe añadir que, las precursoras del proyecto dictaminan en las recomendaciones de este trabajo la elaboración y estandarización de alguna plantilla en esta facultad.

De esta manera, el docente *Yony Alejandro Pastrana Caballero* en el año 2017 desarrolló la plantilla en lenguaje *Latex*, para que, de esta manera se pudiese establecer una pauta en la estandarización de las guías de laboratorio. En la **Figura 14** se puede apreciar el aspecto y la diagramación del modelo, en el cual se desarrollaron las guías posteriores, evidenciando una estructura más sobria y completa en la cual la información respectiva y el paso a paso de esta se puede seguir de manera lógica y ordenada. Esto dictamina una pauta importante en el terreno, debido a que permitió acelerar y canalizar esfuerzos directamente en las fuentes de información de lenguaje *R* y de reforzamiento de conceptos de calidad y estadística que se convertirían en el insumo pertinente en el cual el contenido de las prácticas iba a ser desarrollado.

**Figura 13.** Primera plantilla elaborada para prácticas de laboratorio

	<b>01</b>	<b>REQUISITOS GENERALES PARA LA REALIZACIÓN DE PRACTICAS DE INGENIERÍA DE METODOS Y TIEMPOS EN EL SISTEMA CIM</b>		
	Programa:	<b>Ingeniería Industrial</b>	Materia:	<b>Ing Métodos Y Tiempos</b>
	Elaborado por:	<b>Laura Morales, Karen Arce</b>	Código:	<b>II0204</b>
	Versión:	<b>1</b>	Año:	<b>2016</b>

Requisitos generales prácticas de Ingeniería de métodos y tiempos II0204:

<b>ESTUDIANTE</b>	<b>DOCENTE</b>	<b>MONITOR</b>
<ul style="list-style-type: none"> <li>• Alistamiento previo de la practica (GUÍA LABORATORIO)</li> <li>• Reconocimiento previo del sistema (VIDEO).</li> <li>• Asistir puntualmente con bata blanca y carné Institucional.</li> </ul>	<ul style="list-style-type: none"> <li>• Aviso previo al monitor de la práctica a realizar (2 días anticipación).</li> <li>• Asistencia obligatoria a capacitaciones del sistema CIM MPS-500.</li> <li>• Estudio previo funcionamiento del sistema (VIDEO)</li> <li>• Compartir con los estudiantes la guía de trabajo y el material de estudio antes de realizar la práctica.</li> </ul>	<ul style="list-style-type: none"> <li>• Estar capacitado en el manejo y requerimientos del sistema CIM MPS 500.</li> <li>• Alistamiento previo de materiales de trabajo:                             <ul style="list-style-type: none"> <li>○ Pistones</li> <li>○ Tapas</li> <li>○ resortes</li> <li>○ Materia prima</li> <li>○ Cronómetros</li> <li>○ Mesa de trabajo (Si es el caso de la guía 3)</li> <li>○ Carga de la programación.</li> </ul> </li> <li>• Preparación del sistema para la práctica.</li> <li>• Supervisión continua de la práctica, equipos, y buen uso de estos.</li> </ul>

**ENTREGABLES:**

- Video técnico del sistema (Estudiante, monitor, docente).
- Guía #1 Estudiante
- Guía #1 Docente
- Video de funcionamiento Secuencia #1. (Estudiante, docente)
- Guía #2 Estudiante
- Guía #2 Docente
- Video de funcionamiento Secuencia #2. (Estudiante, docente)
- Guía #3 Estudiante
- Guía #3 Docente
- Video de funcionamiento Secuencia #3. (Estudiante, docente)
- Estándar de proceso (Monitor)

(Arce & Morales, 2016)

**Figura 14.** Estructura y diseño de la plantilla guía

## PRÁCTICA 1: NOMBRE DE LA GUÍA

### 1.1. Competencias a desarrollar

- ✓ Competencia uno
- ✓ Competencia dos
- ✓ Competencia tres

### 1.2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows 7 o posterior .....	1
RoboPro .....	1
Módulos FischerTechnik asignados .....	2
Cable USB A-microUSB .....	1
Adaptador de 9V .....	1
(*)Imán .....	1

! (\*) Cada grupo de trabajo debe traer estos elementos.

### 1.3. Teoría relacionada

Explique en forma clara y concisa los conceptos a estudiar, para los títulos de esta sección emplee letra normal negrita. No olvide realizar la citación bibliográfica según las normas APA vigentes y colocar el nombre a las figuras y tablas como se muestra a continuación.

! Este es un cuadro de alerta. Los cuadros de alerta se emplean para llamar la atención del estudiante sobre algún aspecto en particular, se sugiere no abusar de esta herramienta.

Otra forma de enfatizar texto es así: *Texto enfatizado*

#### 1.3.1 Título de ejemplo

Al redactar la Teoría Relacionada evite segmentar en exceso el texto introduciendo muchos títulos o viñetas. Trate que el texto sea conexo y coherente.

En caso de requerir el uso de texto matemático sobre un renglón, puede usar este ejemplo: El

(Pastrana, 2017)

### **6.2.2 CONTEXTUALIZACIÓN DE COMANDOS DE R CON LOS TEMAS DE CADA GUÍA:**

La labor consecuente de buscar un conjunto de paquetes o comandos que permitieran la realización de los gráficos de control respectivos, fue propiamente un menester en el cual se revisó un amplio material, encontrando una diversidad de librerías que permitieron lograr este objetivo. De todas estas, el paquete escogido fue el *Quality Control Charts* o más conocido por sus siglas como “*qcc*”. Este paquete presenta la ventaja de que las funciones no conllevan un amplio desarrollo de estas, ofreciendo facilidad de comprensión para el usuario debido a su sintaxis intuitiva y de clara recordación.

Cabe añadir que, el uso de un solo paquete para poder trabajar estos gráficos de control brinda la gran ventaja de poder adquirir y dominar todas las aristas que ofrece esta librería, haciendo uso de herramientas y argumentos que se complementan armoniosamente para poder llevar a cabo los respectivos objetivos.

### **6.2.3 PLANIFICACIÓN DE LA METODOLOGÍA Y ACTIVIDADES DE LA GUÍA:**

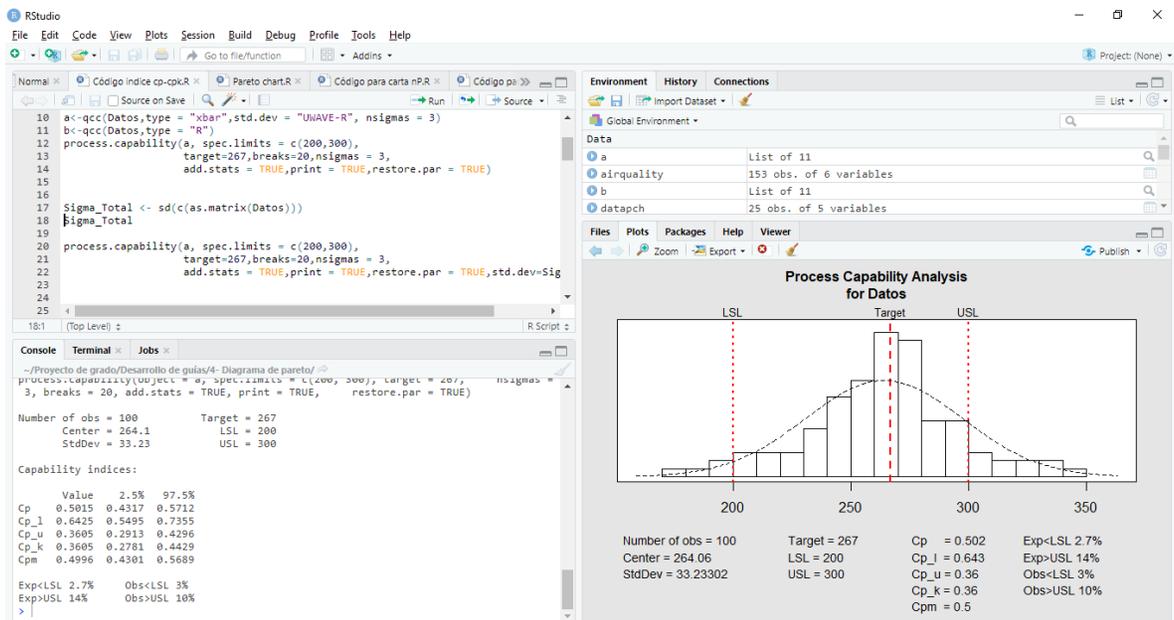
En esta etapa, se realizó la planificación de la estructura informativa, correspondiente al contenido respectivo para las guías de laboratorio. Cada una de ellas aborda los siguientes aspectos de forma ordenada:

- 1. Competencias a desarrollar:** Se establecen las pautas con respecto de los conocimientos y habilidades que el estudiante logrará establecer una vez finalice la práctica.
- 2. Materiales requeridos:** Para todas las guías de laboratorio se deberá disponer de un computador con sistema operativo *Windows*, *Mac OS* o *Linux*; el software de *R* y la interfaz de ambiente integrado (IDE) de *Rstudio*.
- 3. Teoría relacionada:** Un breve panorama a los conceptos a trabajar en la práctica se plasman, con el fin de establecer un contexto claro y conciso al estudiante sobre la temática trabajada.
- 4. Montaje y procedimiento:** Es en este punto donde se establecen las pautas para las líneas de comando a ejecutar en el software, contextualizando los comando a partir de una problemática en concreto. Enseguida, se establecen recomendaciones tanto para el docente como para el estudiante, en lo que respecta al paso a paso respectivo para la satisfactoria realización de la práctica.
- 5. Evaluación:** Una vez se ha finalizado la práctica, se establecen una serie de ejercicios y actividades que deberá desarrollar el estudiante en la entrega del informe de laboratorio.
- 6. Referencias:** Finalmente, se muestran las referencias en las cuales el autor se basó para poder desarrollar el respectivo material.

**6.2.4 DESARROLLO DEL CÓDIGO DE LA GUÍA EN LOS PROGRAMAS CORRESPONDIENTES:** La antesala a la elaboración de la guía iniciaba principalmente, con el desarrollo del código en *Rstudio*. Eso con el fin de verificar y entender la idea y el funcionamiento de los comandos que se desarrollarían *a posteriori* en la guía de laboratorio. De esa manera, se tuvo que realizar varias pruebas y correcciones al código hasta que se pudiesen arreglar los inconvenientes.

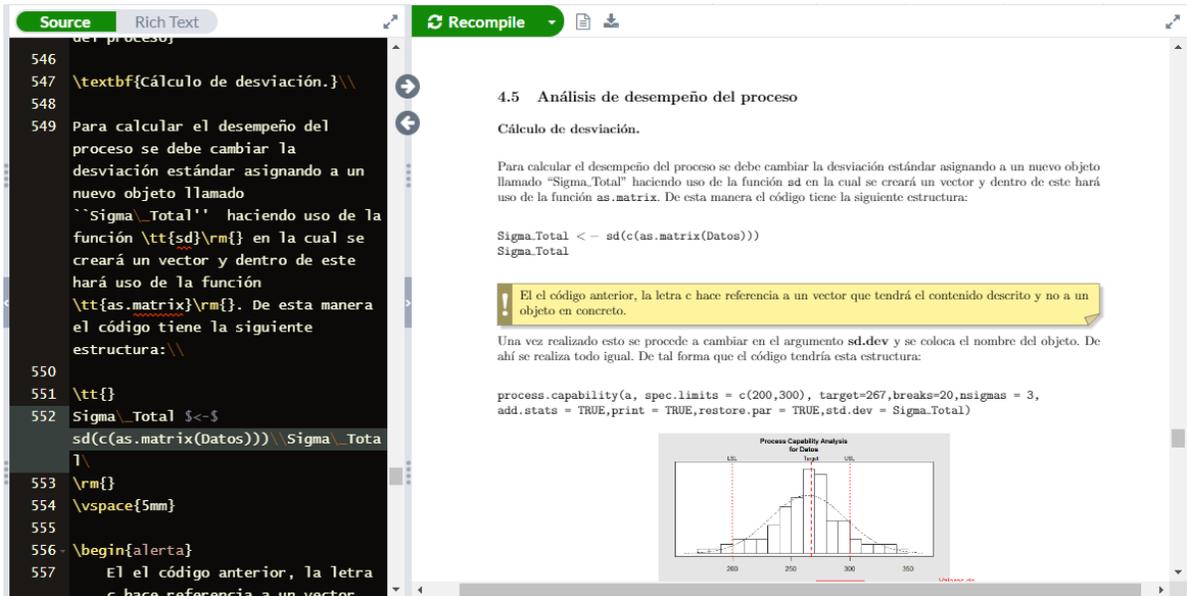
Por otra parte, en la exploración de los argumentos que manejaba cada función, se verificaban y se dictaminaban su correcto funcionamiento ya que, si bien se encontraban sus explicaciones y estructuras en los manuales respectivos para cada paquete, se presentaba el inconveniente de que la información estaba en inglés y que si bien afortunadamente eso no era un impedimento *per se*, había definiciones que resultaban ambiguas y confusas en su entendimiento, con lo cual se tenía la necesidad de investigar en foros de internet especialmente en *Stack Overflow* si algún usuario hubiese presentado ese inconveniente, o en el mejor de los casos, abrir un espacio de discusión elaborando una pregunta en los diferentes foros de *R* que se encuentran en tanto en la cultura anglosajona como en la norteamericana.

**Figura 15.** Puesta en escena de la Interfaz de *Rstudio*



Una vez se establecía la confiabilidad y efectividad del código elaborado, se procedía a plasmarlo en la guía. De manera integral, se hizo un trabajo de elaboración de las guías haciendo uso del lenguaje *Latex*. Se tomó la decisión de trabajar en el editor online de *Overleaf* con asociación a *Sharelatex* por su forma práctica y flexible. Esto permitió que los archivos se pudiesen editar con mucha rapidez y facilidad en cualquier momento y lugar.

Figura 16. Puesta en escena de la Interfaz de Overleaf-Sharelatex.



**6.2.5 DESARROLLO DE LAS INSTRUCCIONES PARA LA ENTREGA DEL INFORME:** Debido a la variabilidad en cuanto a los puntos a desarrollar en cada guía, se dispone a decisión propia del docente de establecer las pautas para la presentación del informe. Con esto se pretende dar mayor margen de flexibilidad al estudiante para poder presentar el desarrollo de los puntos establecidos en la práctica de laboratorio. Cabe decir que, este dictamen fue aplicado por parte del docente *Yony Alejandro Pastrana Caballero*

Figura 17. Métricas de presentación de informe de laboratorio

### INFORME DE LABORATORIO

Indique a sus estudiantes los requisitos para la entrega del informe. No olvide especificar el formato a utilizar, tipo de letra y secciones que debe contener el informe

#### 4.1. ddd

lclcl

(Pastrana, 2017)

## 6.3 FASE DE VERIFICACIÓN Y AJUSTES

**6.3.1 REVISIÓN DEL BORRADOR INICIAL:** Una vez se finalizaba el primer borrador cada guía, esta se sometía a un revisión por parte de la directora del proyecto, lo cual permitió además de corregir las respectivas oportunidades de mejora, también abrir la puerta a una retroalimentación y perspectiva de puntos ciegos que presentaban las guías, en lo que respecta a estructuración y orden del contenido del material correspondiente..

**6.3.2 DESARROLLO DE PRUEBAS:** Una vez se corregían los 13 borradores de las guías, estas eran sometidas a las pruebas correspondientes. La dinámica de los tests consistió generalmente en revisar la teoría respectiva de la guía a elaborar, enseguida, se daba la contextualización del problema a trabajar y se desarrollaba de acuerdo al paso a paso expuesto en el material de aprendizaje el código correspondiente.

Inicialmente, se pretendía disponer del voluntariado de un grupo de estudiantes que tuviesen la disposición de realizar el respectivo material de aprendizaje para a la postre, poder realizar la identificación de las oportunidades de mejora pertinentes que estos presentasen al momento de desarrollar cada guía, sin embargo, la coyuntura mundial se abalanzó de forma implacable.

Figura 18. Estudiante de diplomado realizando prueba

The screenshot shows the RStudio Cloud interface. The main editor window contains the following R code:

```
colors()
boxplot(Solar_8-Month_data=airquality, radiación, main="Medida de radiación",
        xlab="Caja", ylab="Nivel de radiación (Langley)", border = "blue",
        col = "cadetblue", las = 1, plot = TRUE, na.action = na.omit,
        outline=TRUE, horizontal = FALSE, notch=FALSE, ann = TRUE, at=c(1,
        2, 3, 4, 5),
        ,names=c("Mayo", "Junio", "Julio", "Agosto", "Septiembre"),
        varwidth=FALSE, boxex=0.6, staplewex=TRUE, outwex=TRUE, range=0)
```

The console window shows the following error message:

```
Error in boxplot.default(split(mf[[response]], mf[[-response]], drop = drop,
+ )
+ ) : object "radiación" not found
```

The Environment pane shows the following variables:

```
main "Medida de radiación solar"
radiación int [1:153] 290 118 149 313 NA NA 299 99 19...
```

The plot window displays a boxplot titled "Medida de radiación solar" with a y-axis ranging from 0 to 150. The plot shows a distribution of data points with a central box and whiskers.

A video feed of the student, Yuly Andrea Rivera Rincón, is visible in the bottom right corner of the screenshot.

Debido a la pandemia del *Covid-19* el alma mater tuvo que adaptarse a un nuevo ritmo de trabajo en el cual, tanto estudiantes como docentes, acudieron al uso de la conectividad a través de plataformas de comunicación en línea para poder salir campantes. Los estudiantes más interesados en realizar pruebas, fueron aquellos que se encontraban desarrollando el diplomado en *Big Data Analytics*. Sin embargo, pocos participantes de este nicho que tuvieron la fortuna de probar este material. Esto se podría evidenciar como una inexorable desgracia, pero lo cierto es que trajo consigo una oportunidad para poder conectar con personas mucho más allá de nuestro país que evidenciaban un gran interés por ampliar sus conocimientos en este lenguaje de programación estadístico.

**Figura 19.** Participante de la Universidad Nacional sede Palmira en prueba

The screenshot shows the RStudio interface. The script editor contains the following code:

```

1 library(qcc)
2 library(qcc)
3 attach(datos_carta_p)
4 datos_carta_p
5 qcc(Articulos_defectuosos,type = "p", sizes = 300 )
6 nuevo=c(0.013, 0.010, 0.030, 0.020, 0.007, 0.017, 0.010,
7         0.027, 0.000, 0.017, 0.040, 0.013, 0.017, 0.013,
8         0.017, 0.013, 0.030, 0.030, 0.007)/300
9 qcc(Articulos_defectuosos,type = "p", sizes = 300,data.name="antiguos datos",
10      newdata=nuevo, newsizes = 300,newdata.name="datos ciclo PHVA")
11 qcc(Articulos_defectuosos,type = "p", sizes = 300,data.name="antiguos datos",
12      newdata=nuevo, newsizes = 300,newdata.name="datos ciclo PHVA")
13 qcc(nuevo,type = "p", sizes = 300, title = "Carta p para datos ciclo PHVA")
14

```

The console shows the execution of the code, including the command `qcc(data = nuevo, type = "p", sizes = 300, title = "carta p para datos ciclo PHVA")` and its output, which includes a list of statistics and a plot. The plot, titled "Carta p para datos ciclo PHVA", shows a line graph with data points and control limits (UCL, CL, LCL) over 19 groups. The y-axis represents the proportion of defective articles, ranging from 0.00 to 0.04. The x-axis represents the group number, from 1 to 19.

**Figura 20.** Autor realizando prueba con participantes de Ecuador.

The screenshot shows a Google Meet session. The main window displays a presentation slide with the text "Estás presentando para todos" and a "Dejar de presentar" button. The chat window on the right shows a list of participants and their messages:

- Shirley Tenorio Obítas: 13:07 Buena tarde, escucho perfectamente
- Yamiko Espinoza Canto: 13:08 Buena tarde, escucho perfectamente
- Yamiko Espinoza Canto: 13:10 si tengo el programa
- Yamiko Espinoza Canto: 13:10 pero quiero aprender sobre el programa
- Shirley Tenorio Obítas: 13:10 Si tenemos nociones respecto al tema

The chat window also includes a "Enviar un mensaje a todos" button and a "Chat" header.

Gracias a la inclusión del autor por parte del grupo de *Facebook r con Rstudio-MDA*, muchas personas se interesaron por hacer parte de este menester, con lo cual, permitieron que el desarrollo de las pruebas se llevase a cabo. Si bien hubieron participantes de Colombia desde Barranquilla hasta Antioquia, también se presentaron personas de Perú y Ecuador.

**6.3.3 REAJUSTES PERTINENTES:** Una vez finalizadas las pruebas con los participantes, se realizaban los ajustes pertinentes, de acuerdo a lo identificado en cada una de las pruebas. Cabe decir que, principalmente estas oportunidades de mejora no se evidenciaban en el código *per se*, si no en la estructuración del paso a paso elaborado para la explicación de este, ya que muchas veces el participante se presentaba con un inconveniente en un párrafo en concreto cuando *a posteriori* de este se evidenciaba otro párrafo con la explicación de esta problemática.

También se pudo apreciar que algunos participantes presentaban una inadecuación del código al desarrollar las guías en *Rstudio Cloud*, con lo cual, se establecieron pautas en todo el material para poder llevar a buen puerto el contenido cuando se elaborase en esta interfaz en línea.

## 7 CONCLUSIONES

Es un hecho innegable que, la inmersión al aprendizaje en el lenguaje de programación *R* es una tarea amplia, debido a que este cuenta con un apoteósico y vasto número de funciones, comandos y argumentos los cuales hacen de este un estudio bastante diverso y complejo. Esto es un hecho bastante importante, el cual permitió una investigación a cabalidad de todas las aristas que se pueden abordar en los aspectos que *R* abarca, especialmente, dictaminando un panorama y reordenación de todos los paquetes que, en lo que respecta a las aplicaciones en control estadístico de calidad se refiere. Con lo que, debido al auge y demanda de esta relativamente nueva herramienta, se han logrado diseñar paquetes como el “qcc” y otros que cumplen a cabalidad con las necesidades de los requerimientos académicos.

En lo que respecta a la implementación y diseño de una plantilla, se han encaminado esfuerzos desde ya hace un par de años en este menester lo cual, ha hecho que se logre establecer un panorama integral en pro de la estandarización de las estructuras teóricas de las prácticas, en cuanto a asignaturas que no sean estrictamente del nicho de las ciencias básicas, tales como control de calidad y demás se refiere. Esto en contraste a, materias de las ramas de la física y afines las cuales, cuentan hasta con incluso manuales para el desarrollo de sus prácticas respectivas. Por otro lado, la cimentación de una estructura de código para el material de aprendizaje, permite establecer una métrica para mediano y largo plazo, que permita la evolución de los conceptos a trabajar por parte del docente para que, con este fin, se permita una flexibilización de los procesos a elaborar en cada una de las sesiones de clase.

Finalmente, cabe decir que la labor respectiva de las revisiones preliminares a los borradores y las pruebas, junto con los ajustes pertinentes realizados en cada una de estas dos etapas, en pro de garantizar y asegurar el satisfactorio desarrollo de las guías tuvo un impacto tal que, se vislumbra *a posteriori* que este compendio será una herramienta que inexorablemente va a realizar un aporte significativo a los procesos tanto de aprendizaje como de enseñanza, en lo que respecta al auge y tendencia a la cual el mundo moderno se encuentra transitando. Esto se evidenció notoriamente en el desarrollo de las pruebas en donde de forma unánime, los participantes expresaron el total impacto que el material tuvo al momento de aprender sobre lenguaje *R*. Estas personas enfatizaban que, la enseñanza de estos códigos y paquetes por parte de sus docentes, no se realizaba de forma en la cual, se explicara el uso y la aplicación concreta de cada uno de los comandos utilizados. Esto permite que, el estudiante comprenda desde los adentros de los códigos que trabaja, el verdadero uso de estos, con lo cual, se permite empoderarse de una forma mucho más eficiente de la información adquirida.

## 8 RECOMENDACIONES

Se recomienda que, se cuente a futuro con un compendio literario en castellano, tanto en formatos físicos como digitales, mucho más amplio sobre este tema en la institución. Esto debido a que, la gran mayoría de la información consultada sobre *R*, bien sea en forma de manuales y material audiovisual, provienen de la literatura en inglés. Esto es crucial mencionarlo debido a que, aún existen muchas brechas por parte de los estudiantes en lo que respecta a sus conocimientos en esta lengua, lo cual implica que una proporción considerable de estos no cuentan con las habilidades necesarias para poder hacer frente a estos inconvenientes investigativos en el idioma extranjero.

Es importante también, abrir un horizonte con respecto a un nicho inclusive externo a la *Fundación Universitaria Los Libertadores*, el cual permita establecer un canal de comunicación entre otras personas, que hagan parte de aquellos que se encuentren inmersos en este lenguaje de programación con enfoque estadístico debido a que, aproximadamente el 61% de pruebas, se realizaron con personas de otras instituciones, hasta incluso de academias fuera de nuestro país y de no ser así, estas no se hubiesen podido llevar a cabo.

Finalmente, se sugiere que el desarrollo de material de aprendizaje en *R* vaya mucho más allá de lo elaborado en estas guías de laboratorio, abarcando diferentes paquetes y librerías, permitiendo así al alma mater, posicionarse como una institución que cuente con un poderoso arsenal de conocimientos bien estructurados sobre este tipo de lenguaje estadístico. Esto con el fin de dar rienda suelta al avance de un paso importante en el nutrimento de la literatura en habla hispana sobre *R*.

ANEXOS

# **ANEXO A. GUÍAS DE APRENDIZAJE**

En este apartado se plasma el compendio de las guías de aprendizaje elaboradas.

# PRACTICA 1: DIAGRAMA DE TALLO Y HOJA

## 1. Competencias a desarrollar

- ✓ Capacidad de realizar de un diagrama de tallo y hoja en lenguaje *R*.
- ✓ Toma de decisiones apropiadas a partir del comportamiento de los datos.
- ✓ Explorar la forma distribucional de un conjunto de datos.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

### 3.1 Definición.

Un diagrama de tallo y hoja es una herramienta gráfica que permite de una forma sencilla visualizar un conjunto de datos. Este gráfico se dió a conocer por el matemático estadounidense *John Tukey* en su libro *Análisis exploratorio de datos* del año 1977. La creación de esta clase de diagramas es fácil de realizar. Debe seguir estos pasos:

1. Realice dos columnas y nómbrelas como “**Tallos**” y “**Hojas**” respectivamente.
2. En la primera columna coloque el primer dígito (decena) de izquierda a derecha y en la segunda columna debe colocar las unidades que compartan al dígito de decena que se encuentra en la columna “**Tallos**”

En la *Figura 1* se evidencia el uso del diagrama de tallo y hoja en el cual se han ordenado las calificaciones obtenidas por 30 estudiantes en un examen [1]:

**Calificaciones:** 63, 58, 61, 52, 59, 65, 69, 75, 70, 54, 57, 63, 76, 81, 64, 68, 59, 40, 65, 74, 80, 44, 47, 53, 70, 81, 68, 49, 57 y 61.

Tallos	Hojas
4	0479
5	82947937
6	3159348581
7	50640
8	101

Figura 1: Diagrama de tallo y hoja de calificaciones

Las interpretaciones del diagrama de tallo y hoja pueden ser diversas dependiendo del contexto de la muestra o población observada. Sin embargo, a rasgos generales se pueden evidenciar dos comportamientos principales: La gráfica presenta **simetría**, es decir, si las proporciones del gráfico son iguales en ambos lados. Por otro lado se presenta el caso de ser **asimétrica o sesgada** con lo cual se observaría una tendencia hacia los datos mayores o menores bien sea en la parte inferior o superior del diagrama. Además de esto, pueden evidenciarse muchas otras medidas estadísticas y novedades en la información con el comportamiento que están presentando los datos tales como valores atípicos, curtosis o apuntamiento de la distribución, entre otros.

## 4. Montaje y procedimiento

! Durante la realización de la practica, permanezca atento a las indicaciones del docente.

### 4.2 Elaboración manual del diagrama de tallo y hoja en Rstudio

Para el desarrollo de esta práctica de laboratorio se tomarán en cuenta los datos de rendimiento semanal de unas instalaciones de fabricación de semiconductores organizados en la tabla 1:

Tabla 1: Datos de rendimiento de semiconductores [2]

Semana	Rendimiento	Semana	Rendimiento
1	48	21	68
2	53	22	65
3	49	23	73
4	51	24	88
5	51	25	69
6	52	26	69
7	63	27	83
8	60	28	78
9	53	29	81
10	64	30	86
11	59	31	92
12	54	32	75
13	47	33	85
14	49	34	77
15	45	35	82
16	64	36	76
17	79	37	75
18	65	38	91
19	62	39	73
20	60	40	92

**Docente:** Indique a los estudiantes que ejecuten el software *R Studio* y después, creen un vector llamado *semiconductores* el cual estará compuesto de los cuarenta datos de la tabla 1.

**Estudiante:** Realice la creación del vector *semiconductores*. Los vectores en *R* se pueden crear utilizando el comando `c()`, en donde debe ingresar dentro de los paréntesis los datos de la tabla 1 teniendo en cuenta que cada número ingresado debe estar separado de una coma. Una vez finalice deberá asignar este vector al nombre *semiconductores*. Esto se puede hacer fácilmente colocando los símbolos “- >” al final

de vector y escribiendo el respectivo nombre al lado derecho de la flecha. Al final para cargar los datos del vector debe dar click en el botón *Run*, o bien presionar las teclas *Ctrl + Enter*. En la figura 2 se evidencia el proceso correspondiente a la creación del vector:

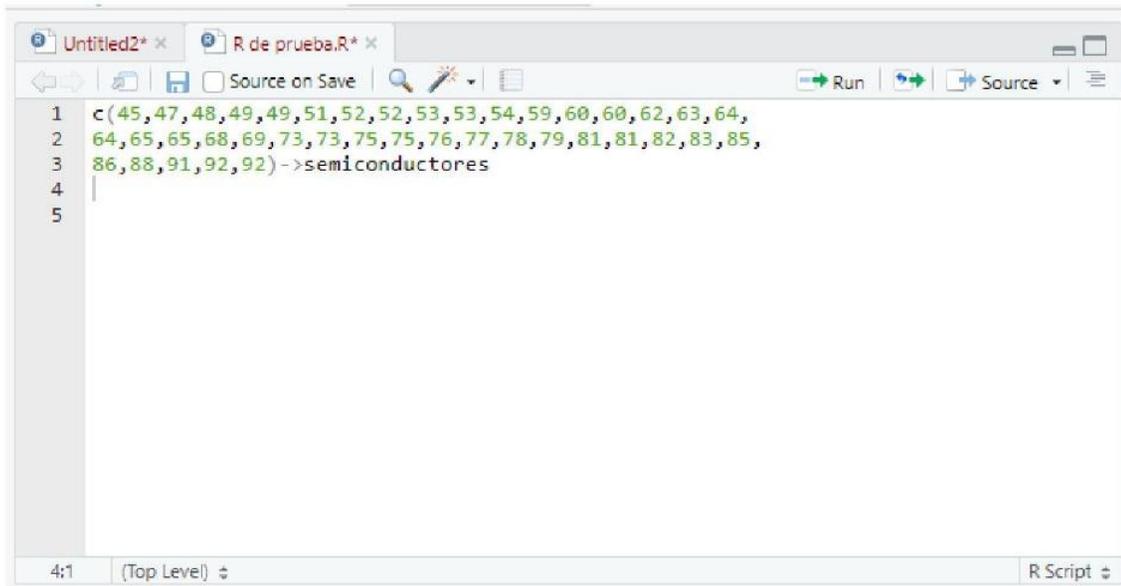


Figura 2: Creación de un vector en el editor de scripts de R

### 4.3 Estructura de la función *stem*

**Docente:** Indique al estudiante el uso de la función *stem* para realizar el diagrama de tallo y hoja en *R* explicando cada uno de los parámetros que se pueden modificar.

**Estudiante:** Utilice la función *stem* de acuerdo a la explicación dada por el docente. Tenga presente que los parámetros que se pueden modificar son:

- ✓ **X:** Es el vector el cual contiene los datos. En este caso será el vector *semiconductores*
- ✓ **Scale:** Sirve para cambiar la longitud del diagrama, su valor por defecto es 1.
- ✓ **Width:** Se utiliza para cambiar el ancho del diagrama, por defecto tiene un valor de 80.
- ✓ **Atom:** Es una tolerancia. De forma predeterminada está en un valor de 1.

Para realizar el diagrama de tallo y hoja simplemente deberá colocar el área de edición de scripts o en la consola la línea: `stem(semiconductores)` . Una vez esto deberá dar click en el botón *Run* o pulsar las teclas *Ctrl + Enter*:

```
Console Terminal x Jobs x
~/Universidad/Proyecto de grado/Desarrollo de guías/1. Diagrama de Tallo y Hoja/ ↗
> stem(semiconductores)

The decimal point is 1 digit(s) to the right of the |

4 | 57899
5 | 122334
5 | 9
6 | 002344
6 | 5589
7 | 33
7 | 556789
8 | 1123
8 | 568
9 | 122

> |
```

Figura 3: Elaboración del diagrama de tallo y hoja en R. Se evidencian que el diagrama cuenta con doble tallo.

*R Studio* por defecto realizará **para una muestra mayor o igual a 25 datos** un diagrama de doble tallo y hoja en la cual el primer tallo agrupará todas las hojas que vayan de 0 a 4, y el segundo tallo realizará el mismo menester para aquellas hojas en un intervalo de 5 a 9. Para cambiar esto se debe escribir la línea: `stem(semiconductores,0.5)` con esto *R Studio* realizará tallos individuales.

```
Console Terminal x Jobs x
~/Universidad/Proyecto de grado/Desarrollo de guías/1. Diagrama de Tallo y Hoja/ ↗
9 | 122

>
> stem(semiconductores,0.5)

The decimal point is 1 digit(s) to the right of the |

4 | 57899
5 | 1223349
6 | 0023445589
7 | 33556789
8 | 1123568
9 | 122

> |
```

Figura 4: Al cambiar el parámetro de scale a 0.5 se ha obtenido un diagrama con tallos individuales

! Si bien los parámetros **Width** y **Atom** son modificables, se recomienda utilizar los valores por defecto ya que esto puede distorsionar el diagrama de tallo y hoja resultante.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica

1. Si cambia el valor de *scale* a 5 ¿Qué sucede con el diagrama?
2. Si cambia el valor de *scale* a 10 ¿Qué sucede con el diagrama?
3. Cambie el valor de *scale* por los valores 50,100,500 y 1000 ¿Qué sucede con el diagrama? ¿El diagrama sigue cumpliendo con su función y presenta información concisa y rápida de analizar?
4. En qué casos consideraría conveniente cambiar el parámetro *scale*
5. De acuerdo al diagrama de tallo y hoja resultante, describa las razones por las que el rendimiento de los semiconductores tiene tanta oscilación en sus porcentajes. También enuncie tres oportunidades de mejora.

## 6. Referencias

- [1] M. B. Math. Stem and leaf. <http://www.mrbartonmaths.com/resources/keystage3/data/>, s.f.
- [2] D. V. Montgomery, Douglas C y Verbeek. *Control estadístico de la calidad*. Limusa Wiley, 2004.

## PRACTICA 2: HISTOGRAMA

### 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos para poder realizar correctamente histogramas en *R*.
- ✓ Capacidad de adaptar el gráfico de acuerdo a las necesidades y al contexto de la información.
- ✓ Conocimiento de los parámetros modificables y la importancia de éstos.
- ✓ Interpretar la información del gráfico resultante.

### 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

### 3. Teoría relacionada

#### 3.1 Definición

Un histograma es una representación gráfica que agrupa datos de una forma muy similar a lo que se hace en un diagrama de tallo y hoja, sin embargo este presenta la información de forma más completa y detallada con lo cual permite hacer un análisis mucho más intensivo de lo que se está presentando.

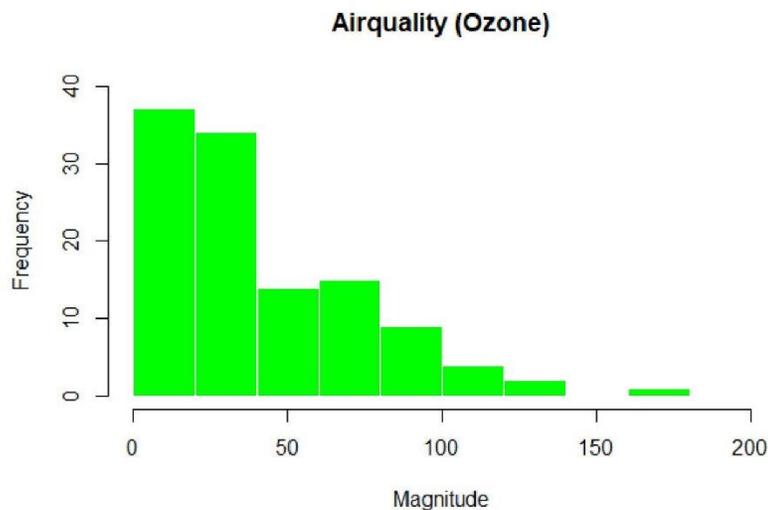


Figura 1: Histograma con los datos de la concentración de ozono que hacen parte del dataset *Airquality*

## 4. Montaje y procedimiento

### 4.2 Creación de histogramas en R studio

En el desarrollo de esta práctica se va a partir del siguiente caso:

“En un restaurante se tiene una fórmula específica para elaborar una cantidad determinada de “agua fresca”, la cual contempla agregar 500 gramos de azúcar. Es claro que resulta de suma importancia añadir exactamente esa cantidad de azúcar para la calidad del agua, ya que, de lo contrario, ésta queda muy dulce o desabrida. Aunque a los cocineros se les ha insistido sobre lo anterior, es frecuente que no pesen el azúcar y la agreguen al tanteo. Al considerar la calidad del agua como un aspecto clave, se decide diseñar un procedimiento a prueba de olvidos: comprar bolsas que contengan 500 gramos de azúcar. Suponga dos marcas de azúcar que cuentan con la presentación de 500 gramos. Al pesarse estas bolsas para cada marca, se obtienen los resultados de la Tabla 1:

Tabla 1: Datos para tamaño de bolsas de azúcar para dos marcas [1]

Marca A	Marca B
503 507 492 499 498 506 502	505 492 502 499 496 499 496
502 506 502 505 493 500 489	495 498 501 504 501 498 498
500 492 500 515 510 502 508	499 495 501 500 497 495 500
499 510 494 503 499 508 513	491 493 507 496 492 499 492
502 515 514 507 510 498 507	501 500 497 500 498 496 494
491 507 502 484 500	497 504 496 500 499

Se plantea desarrollar en esta práctica un histograma en *R* para las bolsas de azúcar de la marca A.

#### Importación de archivos a Rstudio.

**Docente:** Indique al estudiante que importe los datos de la tabla 1 en *Rstudio* partiendo de la elaboración de esta en *Excel* y exportando el archivo resultante del tipo “delimitado por tabulaciones”.

**Estudiante:** Realice la importación de los datos correspondientes a *Rstudio*. Para esto primero deberá elaborar la Tabla 1 en una hoja de cálculo en *Excel* (**puede hacerse una idea de la tabla requerida chequeando en el paso 3 de la Figura 3**). En el momento de guardar el archivo lo deberá hacer como “**Texto (delimitado por tabulaciones)**”. Esto le generará un bloc de notas el cual será el archivo que debe ser importado.

Dirijase al entorno de variables en *Rstudio* y haga click sobre la pestaña “Environment”, y luego sobre “Import Dataset”. Se desplegará una lista en la cual debe hacer click en la opción que cita “From text (readr)...”. Aquí se abrirá el cuadro de texto mostrado en la Figura 3 el cual es una interfaz que le permitirá importar el texto delimitado por tabulaciones realizado anteriormente.

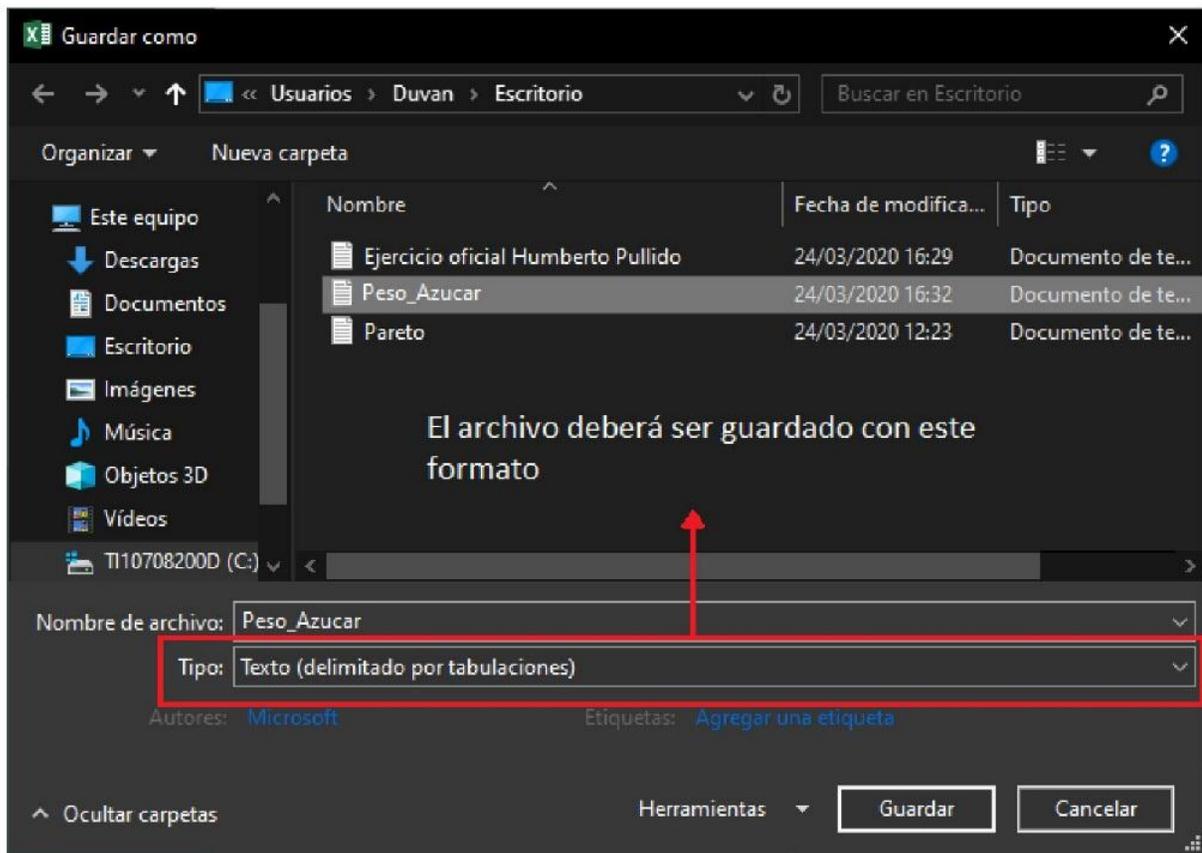


Figura 2: Cuadro de texto en Excel para guardar los datos como texto delimitado por tabulaciones.

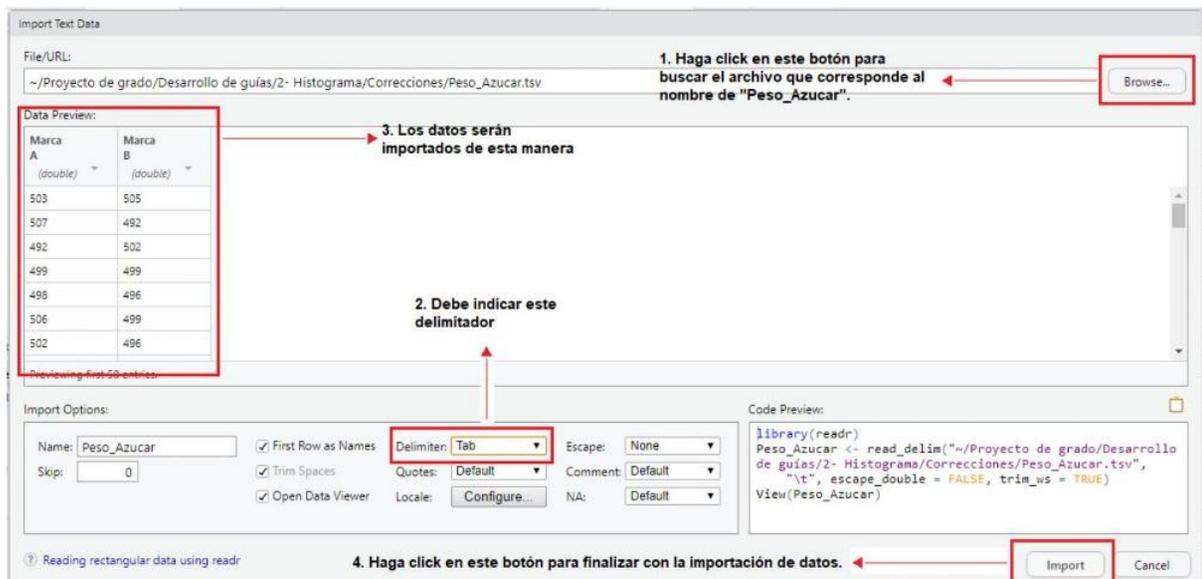


Figura 3: Interfaz para importar datos como texto. Siga los pasos indicados para llevar a cabo la carga de archivos.

## Elaboración de código.

**Docente:** Una vez los datos han sido importados, indique al estudiante que realice el histograma inicial en *Rstudio* con los datos de la marca A.

**Estudiante:** Realice el histograma inicial en *Rstudio*. Para lograr esto deberá utilizar el comando `hist()`. La línea que deberá utilizar se encuentra evidenciada en la figura 4 donde se explica a fondo su estructura básica:



Figura 4: Uso básico del comando `hist()` y su funcionamiento principal

! Tenga presente que el signo `$` tiene la función de llamar solamente los datos de la columna de interés.

En este punto es importante hacer la aclaración sobre el uso del signo `$` en *R* de una manera más pragmática con fines de una clara comprensión.

Si prueba ejecutando la línea `hist(airquality$Ozone)` en el área de edición de scripts o en la consola, *Rstudio* realizará de una forma básica con argumentos por defecto el histograma observado en la *figura 1* ya que, de acuerdo a la información observada en la *Figura 4*, con esta línea estará llamando al dataframe *Airquality* y **con la ayuda del signo `$` podrá traer a escena la columna con los datos deseados que en este breve ejemplo es *Ozone***.

Hay casos en los el usuario desea saber los nombres de las columnas y en general, tener a la mano información del dataset con el que está trabajando. Para saber el nombre de las columnas puede utilizar el comando `names()`. Si por otro lado desea hacerse un pequeño panorama del dataset puede hacer uso de la línea `head()` que le mostrará las seis primeras filas. En la siguiente figura se muestra un breve ejemplo del uso de estos comandos y su resultado en la consola:

```

> names(airquality)
[1] "Ozone" "Solar.R" "Wind" "Temp" "Month" "Day"
> head(airquality)
  Ozone Solar.R Wind Temp Month Day
1    41     190  7.4  67     5    1
2    36     118  8.0  72     5    2
3    12     149 12.6  74     5    3
4    18     313 11.5  62     5    4
5    NA      NA 14.3  56     5    5
6    28      NA 14.9  66     5    6

```

Figura 5: Puesta en escena de las funciones `names()` y `head()` para el dataset *Airquality*.

! *Rstudio* tiene la bondad de guardar dentro de sí una amplia gama de dataframes para su uso, y uno de los más conocidos es *Airquality*. Si desea conocer más sobre los dataframes integrados en *Rstudio* puede consultar el manual *R: A language and Environment for Statistical Computing* o bien pedir indicación a su docente.

Una vez dado este breve paréntesis a estos conceptos, ahora ejecute el código mostrado en la *Figura 4*. De esta manera el histograma resultante tendrá el siguiente aspecto:

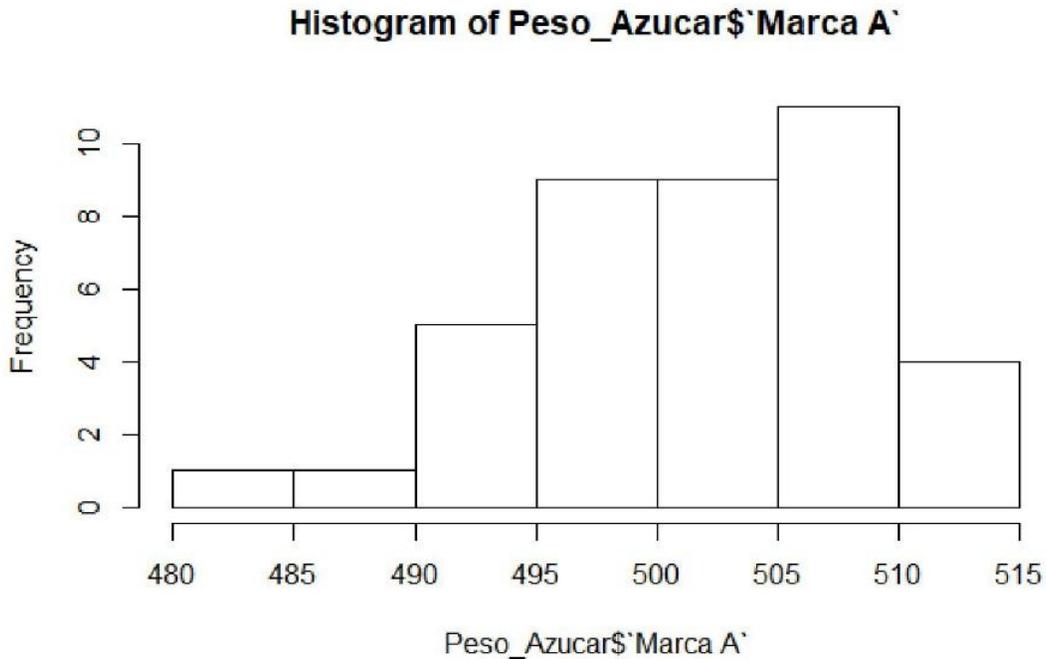


Figura 6: Histograma realizado para las 40 bolsas de la marca A.

#### Argumentos de la función `hist`.

Si bien *R Studio* realiza el gráfico correspondiente, este histograma puede resultar demasiado sobrio para su presentación. Aún así, se pueden realizar varias modificaciones a algunos parámetros de tal manera que esta

herramienta visual cuenta con una presentación mucho más elaborada y además presente mayor información.

**Docente:** Realice la explicación de cada uno de los argumentos y enseguida indique al estudiante que los ejecute paso a paso de acuerdo a los parámetros y valores establecidos en la guía.

**Estudiante:** Modifique los argumentos con los que cuenta el histograma una vez su docente le indique. Los argumentos que se pueden modificar en este tipo de gráficos son:

**!** No olvide separar cada parámetro con una coma de lo contrario el comando no será ejecutado.

- ✓ **x:** Es un vector con los valores especificados, para este caso es la línea del histograma inicial:

```
hist(Peso_Azucar$'Marca A')
```

- ✓ **breaks:** Parámetro numérico o vector. Con este se pueden modificar el número de barras aproximado que desean verse en el histograma. Indique para el gráfico un valor para "breaks" de 5 con la línea:

```
hist(Peso_Azucar$'Marca A', breaks = 5)
```

- ✓ **freq:** Es un argumento lógico; si se coloca "*TRUE*" gráfico presentará sobre las barras las frecuencias absolutas de los datos, en caso contrario, al utilizar "*FALSE*" se colocarán las frecuencias relativas correspondientes. Para este caso se utilizó el argumento "*TRUE*" de la forma:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE)
```

- ✓ **probability:** Es un argumento opcional. Sólo indica un nombre o alias compatible con lenguaje S para "freq". El código que deberá colocar será:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "freq" )
```

- ✓ **include.lowest:** Este argumento lógico añade el valor más bajo al principio del primer intervalo. Al igual que sucede con los diagramas de tallo y hoja, los datos se encuentran agrupados en intervalos abiertos por la izquierda y cerrados por la derecha. Al declarar "*TRUE*" valor del primer intervalo será incluido. Para agregar este parámetro al código actual se deberá escribir:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "freq", include.lowest = TRUE)
```

- ✓ **right:** Argumento lógico el cual dictamina los intervalos abiertos por la izquierda y cerrados por la derecha al declarar "*TRUE*". En caso de declarar "*FALSE*" sucede el caso contrario, es decir, los intervalos serán cerrados por la izquierda y abiertos por la derecha y además, el parámetro "include.lowest=TRUE" tendría el significado de incluir el valor más alto del intervalo. En el histograma a desarrollar este argumento tendrá la forma:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "freq", include.lowest = TRUE, right = TRUE)
```

- ✓ **density:** Permite añadir líneas sobreadas al gráfico y también controlar el número de estas. Para este caso se tomará un valor de 20 y se escribirá de la manera:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20)
```

- ✓ **angle:** Indica el ángulo en grados sexagesimales (de 0 a 360 grados) de las líneas sombreadas. Las líneas contarán con 60 grados colocando en el comando:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20, angle = 60)
```

- ✓ **col:** Sirve para definir el color de las barras del histograma. Se pueden llamar los colores mediante números o directamente colocando el nombre del color en inglés. Hay que tener presente que al haber declarado el argumento **density** las líneas sombreadas serán coloreadas. En el gráfico en cuestión estás tendrán un tono rojizo y para hacerlo se añade a la línea:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20, angle = 60, col = "green")
```

- ✓ **border:** Sirve para definir el color que tendrán el borde de las barras. Se pueden llamar de la misma manera que en el argumento **col**. En este caso las barras tendrán un color negro. Para hacerlo se añade al comando el parámetro:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black")
```

- ✓ **main:** Este argumento se utiliza para dar nombre al gráfico. Al código realizado se le colocará el nombre de "Bolsas de 500 g de la marca A" y se añadirá al comando en cuestión de la forma:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas
      de 500 g de la marca A")
```

- ✓ **xlab - ylab:** Sirve para dar la etiqueta o nombre al eje X e Y respectivamente. En el gráfico se que se está desarrollando, se aprecian en el eje X los grados en escala de Richter, por lo tanto, este eje será nombrado en **xlab** como "Escala de Richter". Por otra parte el eje Y muestra las frecuencias de los eventos, con lo que **ylab** será nombrado como "Frecuencia". La forma para cambiar estos argumentos en el comando respectivo es:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
      500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia")
```

- ✓ **xlim - ylim:** Se utilizan para determinar los intervalos en los que se verán los datos en los ejes. (No confundir con "breaks"). El histograma que se está desarrollando en la práctica deberá contar que en el eje X los límites se encuentren entre 4 hasta 7, y en el eje Y desde 0 hasta 600. Para lograr esto se deben añadir como vectores en el comando principal de la siguiente forma:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
      TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
      500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia", xlim = c(4,7), ylim =
      c(0,600))
```

- ✓ **axes:** Determina la posibilidad de visualizar los ejes del histograma al estar declarado como *TRUE*. Por otro lado, si se declara como *FALSE* sólo se podrán visualizar las barras. Para el gráfico que está elaborando deberá colocar este parámetro como *TRUE* de la siguiente manera:

```
hist(Peso_Azucar$`Marca A`, breaks = 5, freq = TRUE, probability = "lfreq", include.lowest =
TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia", xlim = c(480,520), ylim = c(0,20),
axes = TRUE)
```

- ✓ **plot:** Este argumento lógico es el que genera el gráfico del histograma si se encuentra en estado *TRUE*. Para este caso deberá declararlo como verdadero añadiéndolo al comando de la forma:

```
hist(Peso_Azucar$`Marca A`, breaks = 5, freq = TRUE, probability = "lfreq", include.lowest =
TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia", xlim = c(480,520), ylim = c(0,20),
axes = TRUE, plot = TRUE)
```

En el caso de encontrarse como *FALSE*, R mostrará en la consola información del conteo de los datos como se evidencia en la Figura 2:

```
$breaks
[1] 480 485 490 495 500 505 510 515

$count
[1] 1 1 5 9 9 11 4

$density
[1] 0.005 0.005 0.025 0.045 0.045 0.055 0.020

$mids
[1] 482.5 487.5 492.5 497.5 502.5 507.5 512.5

$xname
[1] "Peso_Azucar$`Marca A`"

$equidist
[1] TRUE

attr(,"class")
[1] "histogram"
```

Figura 7: R realiza un conteo de los parámetros y datos al declarar como *FALSE* el argumento **plot**.

- ✓ **labels:** Argumento lógico que coloca cifras arriba de las barras al ser declarado como verdadero. El gráfico que está elaborando deberá estar declarado como *TRUE* añadiendo al comando:

```
hist(Peso_Azucar$`Marca A`, breaks = 5, freq = TRUE, probability = "lfreq", include.lowest =
TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia", xlim = c(480,520), ylim = c(0,20),
axes = TRUE, plot = TRUE, labels = TRUE)
```

- ✓ **warn.unused:** Argumento lógico. Si se declara como *TRUE* y **plot** como *FALSE*, dará un mensaje de advertencia indicando que el gráfico no está con los parámetros por defecto. En el histograma que está elaborando declare este argumento como verdadero:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia", xlim = c(480,520), ylim = c(0,20),
axes = TRUE, plot = TRUE, labels = TRUE, warn.unused = TRUE)
```

- ✓ **las:** Argumento numérico cuya función es cambiar la orientación de los valores de los ejes. Sólomente puede utilizar los valores 0, 1, 2 y 3. En el histograma pertinente declare este parámetro con un valor de 1. Con esto, todos los valores tanto del eje X como del eje Y se verán de forma vertical:

```
hist(Peso_Azucar$'Marca A', breaks = 5, freq = TRUE, probability = "!freq", include.lowest =
TRUE, right = TRUE, density = 20, angle = 60, col = "green", border = "black", main = "Bolsas de
500 g de la marca A", xlab = "Gramos (g)", ylab = "Frecuencia", xlim = c(480,520), ylim = c(0,20),
axes = TRUE, plot = TRUE, labels = TRUE, warn.unused = TRUE, las = 1)
```

- ✓ **nclass:** Argumento compatible con lenguaje *S*. Realiza la misma función que **breaks**. En este caso no debe ser añadido ya que se declaró en el principio que el parámetro **breaks** contara con un valor de 5.

El histograma una vez realizadas todas las modificaciones correspondientes a sus respectivos argumentos debe tener un aspecto tal como se evidencia en la figura 8:

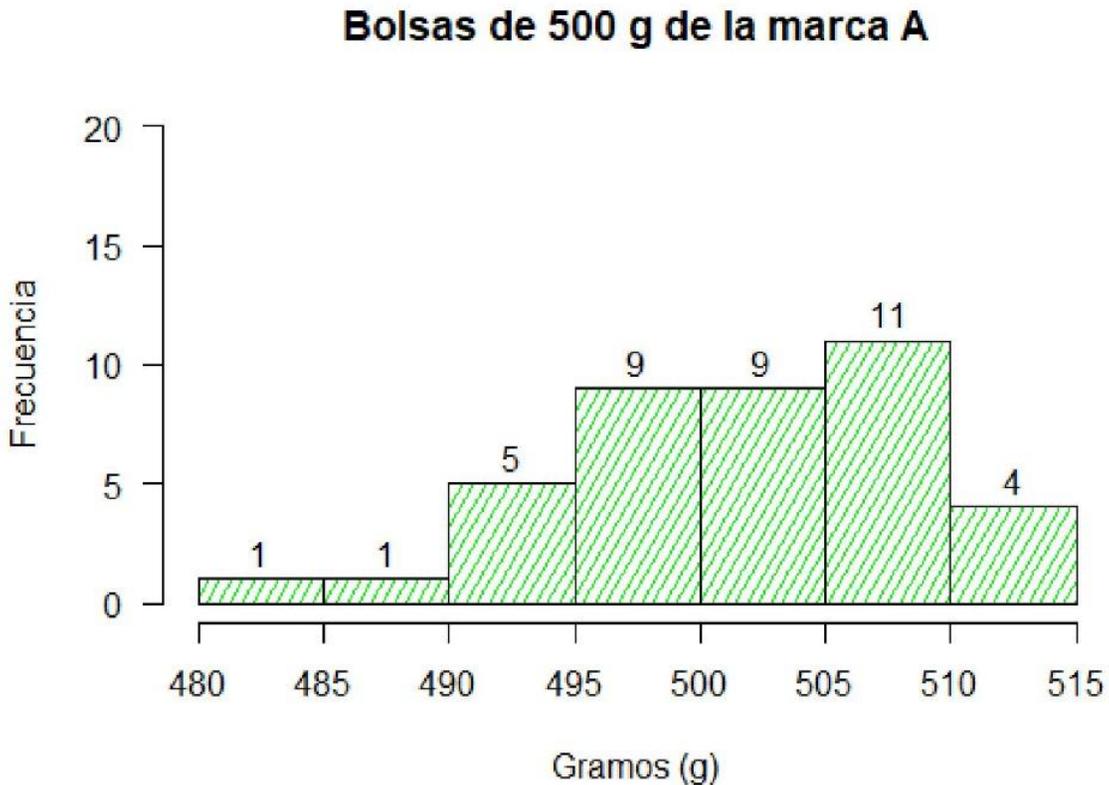


Figura 8: Histograma una vez se han realizado las modificaciones correspondientes

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

1. Recuerde que la función del argumento **breaks** es la de cambiar el número aproximado de barras visibles en el histograma que desea desarrollar. En ese orden de ideas, cambie la estructura del argumento por la siguiente línea:

```
hist(Peso_Azucar$`Marca A`,breaks = seq(480,520, by = 2 )
```

- ¿Qué cambio sufrió el gráfico? ¿qué ventajas tiene declarar de esta manera a este argumento?.
2. ¿Considera pertinente que el argumento **warn.unused** (utilizado para indicar que los parámetros no están por defecto cuando el argumento **plot** es falso) deba ser modificado o es mas conveniente que se deje por defecto declarado como verdadero?
3. Cambie el orden de los argumentos en el código que ha desarrollado: ¿Que sucede? ¿Existe algún inconveniente o error?¿Qué puede concluir al respecto?

## 6. Referencias

[1] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010.

- Nefli Correa, Juan C y Gonzalez. *Gráficos estadísticos con R*. Universidad Nacional de Colombia Sede Medellín, 2002

- Karlijn Willems. How to make histograms with basic r. <https://www.datacamp.com/community/tutorials/make-histogram-basic-r>, 2019.

## PRACTICA 3: DIAGRAMA DE CAJA

### 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos para poder realizar correctamente diagramas de caja en *R*.
- ✓ Capacidad de adaptar el gráfico de acuerdo a las necesidades y al contexto de la información.
- ✓ Conocimiento de los parámetros modificables y la importancia de éstos.
- ✓ Interpretar la información del gráfico resultante.

### 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

### 3. Teoría relacionada

#### 3.1 Definición.

Un diagrama de caja o también conocido como diagrama de caja y bigote, es una representación que permite evidenciar de una forma sencilla la distribución de los datos acudiendo a sus cuartiles. Esta herramienta visual consta de una caja en la cual la parte inferior pertenece al primer cuartil, la línea central corresponde a la mediana o segundo cuartil, y el borde superior de la caja representa al tercer cuartil. También se dibujan unos límites para los valores máximos y mínimos de la población estos datos son los llamados “bigotes”. Hay casos en los que se pueden apreciar datos más allá de estos “bigotes” los cuales son valores atípicos de la muestra o la población evaluada.

#### Paso a paso para hacer diagramas de caja.

El proceso de elaboración de un diagrama de caja consta de los siguientes pasos:

1. Ordenar el conjunto de datos de menor a mayor.
2. Obtener los cinco puntos fundamentales para la elaboración del gráfico.
  - ✓ Identificar los valores máximos y mínimos del conjunto de datos.
  - ✓ Obtener la mediana correspondiente. Recuerde que si el número valores es impar, esta medida es justamente la mitad de los datos. Por otro lado, si el número de valores es par, será la media de los dos números que se encuentran en la mitad justamente.
  - ✓ Obtener los cuartiles del conjunto de datos. Si bien no hay un consenso en cuanto a un sólo método para hallar estas medidas de posición, una ecuación efectiva para este menester es:

$$L = \frac{k}{100} \cdot n[1]$$

Donde  $L$  es el valor del percentil que se desea obtener;  $k$  es el número de ubicación del percentil (recuerde que los percentiles son medidas de posición que dividen el conjunto de datos en 99 puntos. En ese orden de ideas, los cuartiles corresponden a 25, 50 y 75). Finalmente,  $n$  es el número de valores del conjunto de datos.

3. Dibujar la caja correspondiente basándose en una escala de medida, ubicando los cinco puntos encontrados en el inciso anterior realizando una caja con los cuartiles y unas extensiones hasta los valores máximos y mínimos tal como se aprecia en la *figura 1*:

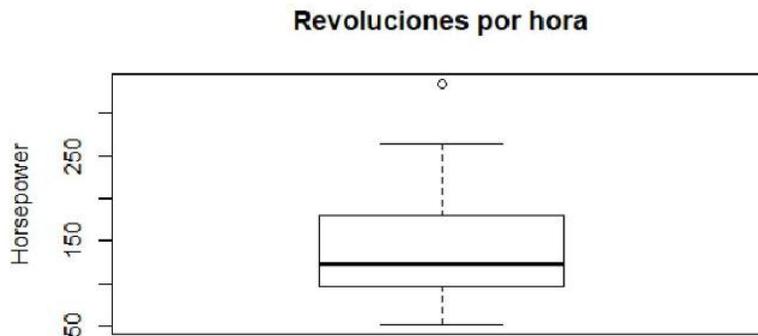


Figura 1: Diagrama de caja correspondiente a caballos de fuerza de 32 automóviles del dataset Mtcars.

## 4. Montaje y procedimiento

### 4.2 Elaboración de diagramas de caja en *Rstudio*.

En esta práctica se tomarán los datos del dataset **airquality** los cuales dictaminan mediciones diarias de la calidad del aire en la ciudad de Nueva York desde mayo hasta septiembre del año 1973.

**Docente:** Indique al estudiante que abra el software *R studio* y que realice un diagrama de caja con los datos del dataset **airquality** con la magnitud **Solar.R**. Luego dé la instrucción para realizar las modificaciones una por una a los argumentos pertinentes.

**Estudiante:** Llame el dataset correspondiente y utilice el comando **boxplot**. La forma general de esta función es:

! Cuando se habla de la forma general de una función, esto sólo indica la forma correcta de su uso y el tipo de argumentos que esta utiliza (lógico, numérico, vector, entre otros). **No se debe ejecutar esta línea.**

```
boxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE,
notch = FALSE, outline = TRUE, names, plot = TRUE,
border = par("fg"), col = NULL, log = ,
```

```
pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5),
ann = !add, horizontal = FALSE, add = FALSE, at = NULL)
```

Por otra parte, utilice la línea `names()` para que conozca todas las magnitudes del dataset **airquality**. Una vez realice esto, deberá permanecer atento a las indicaciones del docente el cual le dará pautas para que uno por uno vaya modificando y conociendo los argumentos correspondientes para un diagrama de caja.

! Añada de forma ordenada cada argumento correspondiente a esta función.

### Elaboración de código.

El código correspondiente para la elaboración del primer gráfico de control se evidencia a continuación:

```
radiación <- airquality$Solar.R
boxplot(radiación,main="Medida de radiación solar",
xlab="Caja",ylab="Nivel de radiación (Landgleys)",
border = "blue",col = "cadetblue1",las=1,plot=TRUE)
```

Los diagramas de caja tienen algunos argumentos en común con los histogramas, es decir, se pueden utilizar de la misma forma para ambos gráficos. Estos argumentos son:

- ✓ **x:** En este caso será el vector para los datos de la radiación solar. Como solamente se trabajará con los la columna **Solar.R** que indica los niveles de radiación del dataset, deberá asignar a `airquality$Solar.R` el nombre de "radiación". Esto se realiza colocando los símbolos "< -" al lado derecho del nombre que desea dar al conjunto de datos:

```
radiación <-airquality$Solar.R
```

Una vez hecho esto, puede empezar a hacer uso de la función `plot` con este argumento. De esta manera el código en cuestión deberá ser declarado de la forma `boxplot(radiación)`

- ✓ **main:** Coloca el título al gráfico. Declare `main='Medida de radiación solar'`
- ✓ **xlab, ylab:** Este argumento cambia las etiquetas de los ejes x e y respectivamente. Dictamine de la forma `xlab="Caja", ylab="Nivel de radiación (Landgleys)'` respectivamente.
- ✓ **border:** Cambia el color de los bordes del diagram de caja. Utilice un color azul rey con lo cual deberá declarar `border="blue"` este argumento.
- ✓ **col:** Especifica el color de relleno del gráfico. Se utilizará un azul turquesa por lo cual deberá declarar `col="cadetblue4"` para este gráfico.
- ✓ **las:** Argumento numérico cuya función es cambiar la orientación de los valores de los ejes. Solamente puede utilizar los valores 0, 1, 2 y 3. Utilice un valor de 1 de la forma `las=1` para que los valores se vean de forma vertical.

- ✓ **plot:** Este argumento lógico es el que genera el gráfico si se encuentra en estado “*TRUE*”. Declare `plot=TRUE` para poder visualizar el gráfico. De lo contrario se le mostrarán datos en la consola del dataset escogido.

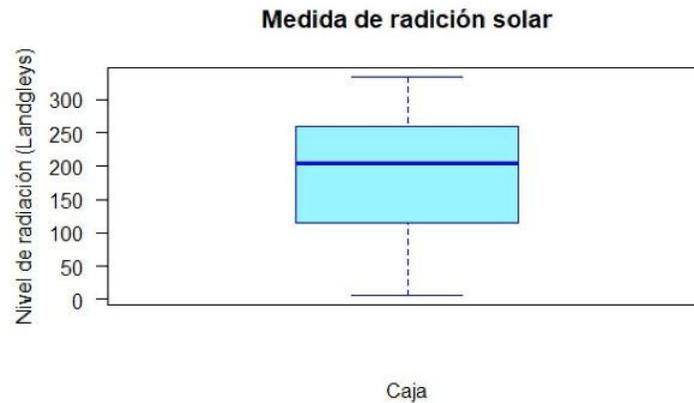


Figura 2: Diagrama de caja inicial

#### Argumentos de la función `boxplot`.

En la *Figura 2* se aprecia el resultado del primer arreglo. Los demás argumentos a tener en cuenta al momento de elaborar un diagrama se encuentran en la siguiente línea de código que cuenta con los parámetros trabajados en el inciso anterior. En cada argumento se explica el paso a paso de todos los argumentos utilizados en orden:

```
boxplot(Solar.R Month,data=airquality,radiación,main="Medida de radiación solar",
xlab="Caja",ylab="Nivel de radiación (Landgleys)",border = "blue", col = "cadetblue1",
las = 1,plot = TRUE,ann = TRUE,na.action = na.omit,range=0,varwidth=FALSE,notch=FALSE,
outline=TRUE,boxwex=0.6,,staplewex=TRUE,outwex=TRUE,horizontal = FALSE,at=c(5,6,7,8,9),
names=c("Mayo","Junio","Julio","Agosto","Septiembre"))
```

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escribalas en *Rstudio*.

- ✓ **formula:** Indica una correlación de dos magnitudes, es decir, la primera contiene un vector numérico y la segunda realiza una agrupación de acuerdo al parámetro establecido. Quizás suene ambiguo, sin embargo se puede ver más claro en su aplicación. En el código elimine el parámetro `x` y escriba `boxplot(Solar.R~Month)` Con esto se le indicará a *R* que se desea realizar un diagrama de caja de la radiación solar para los meses en los que se tomaron las respectivas mediciones. los meses estarán registrados de forma numérica (1 es enero, 2 febrero, entre otros).
- ✓ **data:** Corresponde al dataset del cual se van a tomar los datos y del cual se apalanca el argumento anterior (en el caso de ser declarado). Para este caso deberá declarar `data=airquality`.

! Para mantener el código en orden escriba este argumento justo al lado del parámetro donde escribió `boxplot(Solar.R~Month)`, luego puede seguir escribiendo los demás argumentos después del parámetro `plot`.

- ✓ **subset:** Es un argumento opcional. Llama a otro conjunto de datos que se encuentren relacionados tal como con `data` y `formula`. En este caso no es necesario utilizar este argumento.
  - ✓ **ann:** Argumento lógico que indica si los nombres de `xlab`, `ylab` y `main` deben ser mostrados en el gráfico. Declare en este caso `ann=TRUE` para especificar como verdadero este argumento.
  - ✓ **na.action:** Existen datasets en los cuales hay atributos catalogados como “no aplica” (`na`) los cuales no aportan mucha información. Se pueden declarar cuatro tipos de este argumento:
    - `na.omit`: Omite los valores “na” que se encuentren en el dataset.
    - `na.fail`: Si no se encuentran valores de “na” no realiza ninguna modificación. En caso contrario, muestra un error en la consola indicando que no se encuentran precisamente este tipo de valores.
    - `na.exclude`: Elimina los valores del tipo “na” que se encuentren en el dataset.
    - `na.pass`: Pasa por alto los valores de “na”.
- Para este caso el parámetro se dejará declarado como `na.action = na.omit`
- ✓ **range:** Cambia la distancia del rasgo marcado como bigote en la caja pertinente. Si se coloca cero de la forma `range=0` estás serán los valores extremos sin importar que estos sean atípicos. Aparte de esto no permite valores negativos. Por otro lado, el valor máximo es 1.5 veces el rango intercuartílico. En este ejercicio deje este valor por defecto.
  - ✓ **width:** Vector que cambia el ancho de las cajas. Este vector será dejado por defecto.
  - ✓ **varwidth:** Dibuja las cajas y bigotes con una proporción igual a la raíz cuadrada del número de la muestra para cada grupo a medir. Para este caso declare verdadero este argumento como `varwidth=TRUE` debido a que no se evidenciará un cambio sustancial en el gráfico.
  - ✓ **notch:** Al declararse verdadero añade una muesca o solapamiento en forma de “v” que parte desde la línea demarcada como segundo cuartil o mediana. En el diagrama pertinente pruebe declarando verdadero para observar la apariencia del gráfico, sin embargo tendrá que dejarla después como `notch=FALSE` debido a que al modificar algunos argumentos posteriores puede salir un error de conflicto en la consola.
  - ✓ **outline:** Dibuja los valores atípicos cuando está declarada como verdadero. Se colocará `outline=TRUE`.
  - ✓ **boxwex:** Argumento numérico que permite modificar el ancho de las cajas. Si bien se puede marcar cualquier valor, se recomienda usar números decimales para tener un mayor control de este parámetro. En este caso debe utilizar un ancho de 0.6, de tal forma que debe utilizar la línea `boxwex=0.6`.
  - ✓ **staplewex:** Su función es hacer que las líneas de los “bigotes” tengan el mismo tamaño a la línea paralela de la que se encuentra a la caja. Si se argumenta como `FALSE` estas líneas desaparecen. Declárela como verdadera de la forma `slaplewex=TRUE`
  - ✓ **outwex:** Teóricamente cambia el ancho de la línea de valores atípicos a una proporción del tamaño de la caja. Para este ejercicio no será utilizado.

- ✓ **log:** Este argumento indica a *R* si se desea realizar en una escala logarítmica el eje “x” o “y” . En este ejercicio **no será utilizado** sin embargo, la forma para aplicarlo es añadiendo: `log=“x”` o `log=“y”`. o bien para ambos ejes de la forma `log=“xy”`.
- ✓ **horizontal:** Dictamina la dirección de las cajas. Al ser declarada verdadera éstas se verán de forma horizontal. Al igual que realizó con **range** puede probar al declarar este argumento como verdadero, sin embargo en este caso la apariencia del gráfico no es muy estética en esa orientación, por lo que deberá declarar este argumento como `horizontal = FALSE` para el gráfico en cuestión.
- ✓ **add:** Argumento lógico opcional. Si es verdadero añade un nuevo diagrama de caja y bigotes al gráfico actual. Este no será declarado.
- ✓ **at:** Vector que determina donde estarán ubicados las cajas. Se deberá utilizar un vector con los siguientes valores de la forma `at=c(5,6,7,8,9)`
- ✓ **names:** Vector que sirve para nombrar las ubicaciones de las cajas. Se creará un vector que cuente con los meses desde mayo hasta septiembre de la forma: `names = c(“Mayo”, “Junio”, “Julio”, “Agosto”, “Septiembre”)`

De esta manera el gráfico resultante debe tener una apariencia como la siguiente figura:

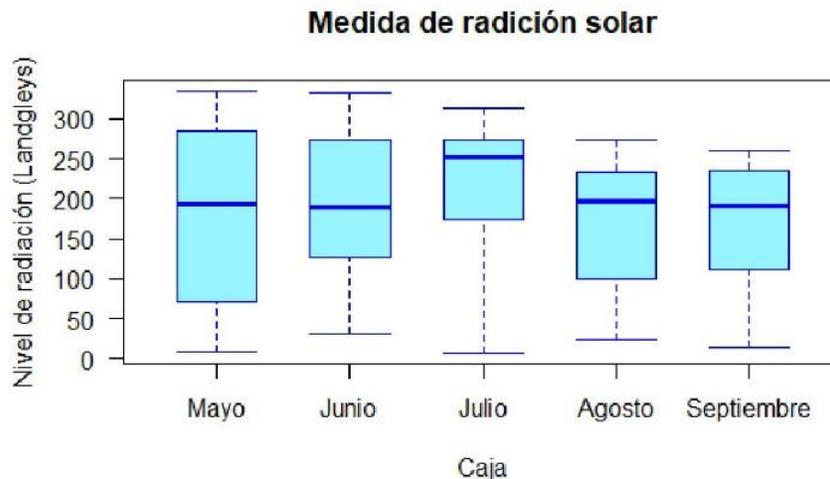


Figura 3: Gráfico final con los argumentos establecidos

! los parámetros **drop**, **sep**, y **lex.order** hacen parte de la función **Split** que permite dividir un vector en subgrupos determinados por otro. Por otro lado, **pars** añade una lista de varios argumentos de la línea **bxp**. Estas funciones son opcionales y no añaden información trascendental.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuesta deben estar relacionadas con lo observado durante la práctica

1. Modifique la escala del argumento **boxwex** por el vector  $c(0.2, 0.5, 0.3, 0.7, 0.8)$  ¿Qué sucede con las cajas?
2. Cambie el orden de los números en el argumento **at** al vector  $c(9, 6, 8, 7, 5)$ . Enseguida, vuelva a **boxwex** y modifique los valores del vector a su voluntad ¿Qué modificaciones es posible controlar con esto?
3. De acuerdo a la información obtenida en el diagrama de caja ¿cúal considera usted que debe ser el comportamiento indicado para los niveles de radiación?. ¿Se deben presentar fluctuaciones tan grandes como en el mes de mayo o se debe mantener como en los meses de agosto y septiembre?

## 6. Referencias

- [1] Triola, Mario. *Estadística*. Pearson Educación, 2004.
- [2] Nefi Correa, Juan C y Gonzalez. *Gráficos estadísticos con R*. Universidad Nacional de Colombia Sede Medellín, 2002.

## PRACTICA 4: DIAGRAMA DE PARETO

### 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente diagramas de pareto en *R*.
- ✓ Desarrollar conocimientos básicos de manejos de librerías en *R*.
- ✓ Conocimiento de los argumentos modificables y la importancia de éstos.
- ✓ Interpretar la información del gráfico resultante.

### 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

### 3. Teoría relacionada

#### 3.1 Definición.

Un diagrama de pareto es una herramienta gráfica que permite realizar un análisis exhaustivo en el que se dictamina el enfoque hacia los problemas vitales, ya que, es imposible abordar todos los problemas al mismo tiempo. Se basa en el *Principio de Pareto* creado por el economista *Vilfredo Pareto* el cual establece que el 80% de los resultados son producidos por el 20% de las causas.

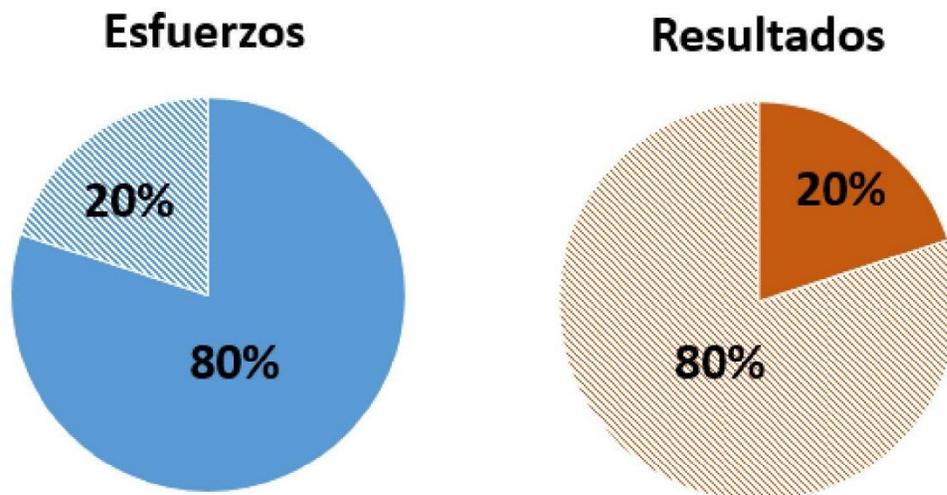


Figura 1: Representación gráfica del principio de Pareto.

## 4. Montaje y procedimiento

### 4.2 Instalación del paquete *qcc*.

**Docente:** Indique a los estudiantes los métodos para la instalación de paquetes en *Rstudio* y la forma de llamarlos para su uso posterior.

**Estudiante:** Siga las indicaciones dadas por el docente en todo lo concerniente al uso de paquetes.

La antesala para la creación de diagramas de pareto en *Rstudio* es la instalación del paquete *Quality Control Chart* abreviado como *qcc*. Esto es un paso fundamental ya que este paquete brinda el entorno necesario para la correcta creación de esta clase de gráficos. Para poder instalarlo existen dos métodos:

1. Ejecute en el área de edición de scripts el comando `install.package()` y en el área entre paréntesis escriba entre comilla el nombre del paquete, de tal forma de que la línea de código debe tener la forma `install.package("qcc")`. Con esto, *Rstudio* pondrá en marcha el comando que le permitirá buscar el paquete en el repositorio de *CRAN* y realizará la respectiva instalación. Si el paquete ha sido instalado correctamente debe salir en la consola la línea de texto que se observa en la *Figura 2*:

```
> install.packages("qcc")
WARNING: Rtools is required to build R packages but is not currently installed. Please
download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/catal/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/qcc_2.7.zip'
Content type 'application/zip' length 3558989 bytes (3.4 MB)
downloaded 3.4 MB

package 'qcc' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
.
C:\Users\catal\AppData\Local\Temp\RtmpojmDKc\downloaded_packages
```

Figura 2: Script ejecutado al momento de la instalación del paquete *qcc*

2. Puede apoyarse en el entorno de utilidades para realizar la instalación. Para esto haga click izquierdo en la pestaña “Packages” y enseguida haga click en el botón “Install”. Con esto se abrirá el cuadro de texto que se muestra en la *Figura 3*:

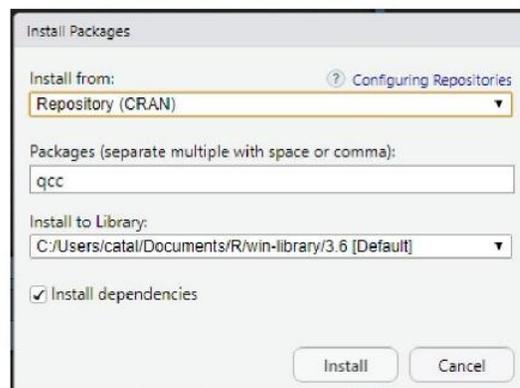


Figura 3: Cuadro de texto correspondiente a la instalación de paquetes.



Tabla 2: Datos para defectos en las piezas en una empresa de manufactura. [2]

Tipo de defecto en las piezas	Número de defectos
R. Superficiales	119
Ruptura	37
Incompletas	13
F. Inapropiadas	12
Otros	9

### Importación de archivos a *Rstudio*.

**Docente:** Indique al estudiante que que realice un diagrama de pareto con los datos de la *Tabla 2* dando la pauta para la creación de variables etiquetadas y realizando las modificaciones una por una a los argumentos pertinentes.

**Estudiante:** Cargue los datos correspondientes a *Rstudio*. Para esto primero deberá replicar la *Tabla 2* en una hoja de calculo en *Excel*. En el momento de guardar el archivo lo deberá hacer como “Texto (delimitado por tabulaciones)”. Esto le generará un bloc de notas el cual será el archivo que debe ser importado. El archivo tendrá como nombre “Ejercicio4”.

! Si está trabajando en *Rstudio Cloud* suba el archivo directamente como un archivo de *Excel*. En caso contrario, siga los procedimientos evidenciados en las *Figuras 5 y 6*.

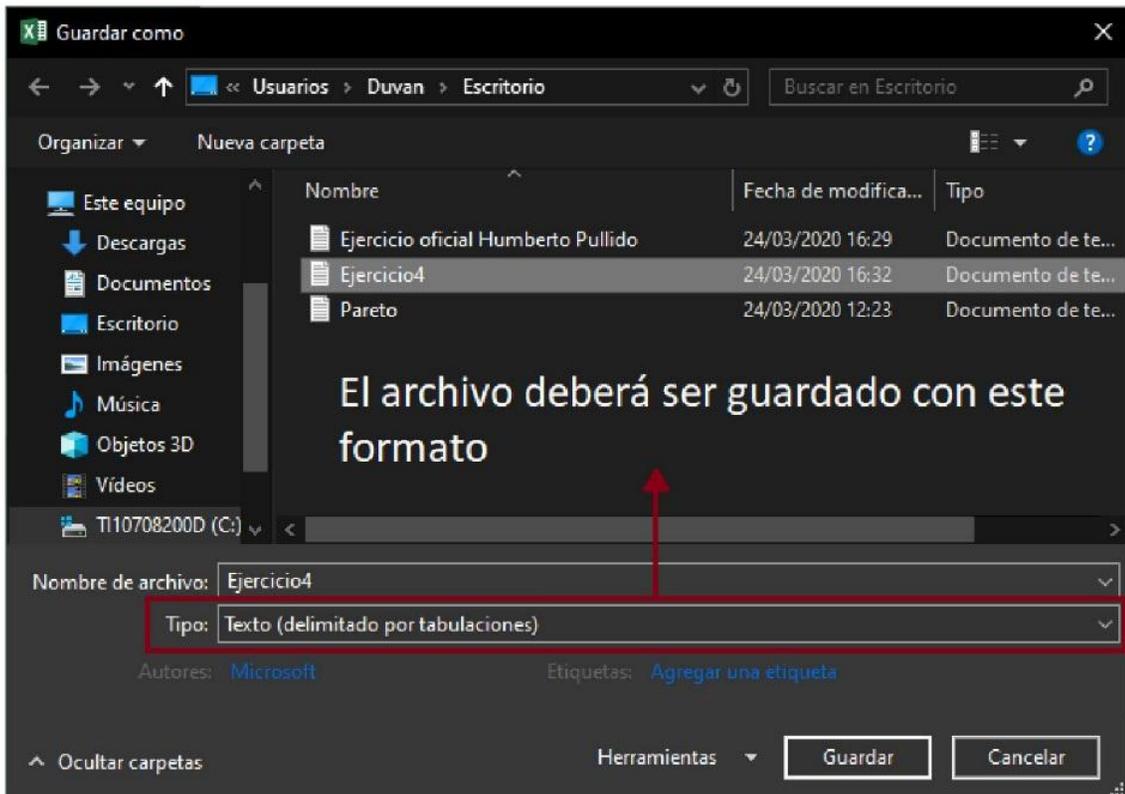


Figura 5: Cuadro de texto en *Excel* para guardar los datos como texto delimitado por tabulaciones.

Diríjase al entorno de variables en *Rstudio* y haga click sobre la pestaña “Environment”, luego sobre “Import Dataset”. Se desplegará una lista en la cual debe hacer click en la opción que cita “From text (readr)...”. Aquí se abrirá el cuadro de texto mostrado en la *Figura 6* el cual es una interfaz que le permitirá importar el texto delimitado por tabulaciones realizado anteriormente:



Figura 6: Interfaz para importar datos como texto. Siga los pasos indicados para llevar a cabo la carga de archivos.

### Elaboración del código.

A continuación, deberá ejecutar la siguiente línea de comando en el área de edición de scripts. Cada una de estas líneas serán explicadas a continuación:

```
attach(Ejercicio4)
names(Ejercicio4)
Defectos<-‘Número de defectos’
names(Defectos)<-‘Tipo de defecto en las piezas’
Defectos
pareto.chart(Defectos)
```

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

1. Llame a la base de datos importada con el comando `attach()` escribiendo el nombre de la base dentro del paréntesis. Con este comando ya no deberá llamar una columna de la forma `Base$Columna` si no que escribirá el nombre de la columna directamente, lo cual le hará más fácil la elaboración del código.
2. Utilice la función `names(Ejercicio4)` para identificar los nombres de las columnas. En este caso serán respectivamente “Tipo de defecto en las piezas” y “Número de defectos”.

3. Asigne el nombre de “Defectos” a los valores que se encuentran en la columna “Número de defectos”. Recuerde que esto se logra utilizando los signos “< -”. De esta manera la línea quedará de la forma `Defectos<-‘Número de defectos’`
4. En este punto se crearán variables etiquetadas. Estas simplemente son parámetros que tienen asignado otro valor (como por ejemplo para este caso, el número de defectos en cada tipo de defecto) Para esto ejecute la línea:

```
names(Defectos)<-‘Tipo de defecto en las piezas’
```

Con esto se le está indicando a *R* que asigne los nombres de la columna indicada con los signos “< -” al conjunto de valores nombrados como “Defectos”.

5. Si se ejecuta solamente la palabra “Defectos”, en la consola se evidenciarán los datos etiquetados.
6. Finalmente, el comando correspondiente para elaborar diagramas de pareto es `pareto.chart()`. La forma general de esta función es:

```
pareto.chart(data, plot=TRUE, xlab, ylab, main, col, las, ylab2, cumperc)
```

en el cual escribirá dentro de los signos de paréntesis el nombre asignado a las variables etiquetadas, en este caso `pareto.chart(Defectos)`. Con esto el gráfico generado tendrá la siguiente forma:

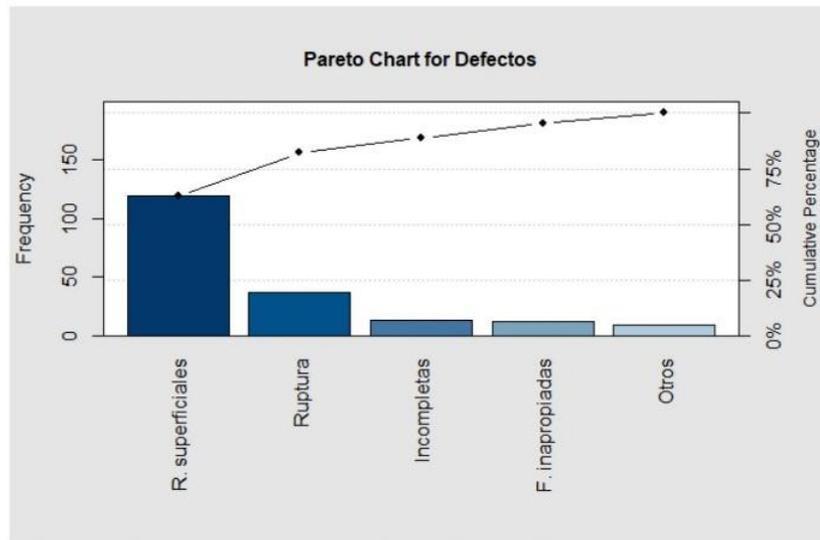


Figura 7: Diagrama de pareto generado para los datos de la variable etiquetada “Defectos”.

Si observa la consola, se habrá ejecutado en orden las siguientes líneas de código:

```

> attach(Ejercicio4)
The following objects are masked from Ejercicio4 (pos = 3):

  Número de defecto, Tipo de defecto en las piezas

The following objects are masked from Ejercicio4 (pos = 4):

  Número de defecto, Tipo de defecto en las piezas

> names(Ejercicio4)
[1] "Tipo de defecto en las piezas" "Número de defecto"
> Defectos<-`Número de defecto`
> names(Defectos)<-`Tipo de defecto en las piezas`
> Defectos
R. superficiales      Ruptura      Incompletas  F. inapropiadas
      119             37             13             12
      Otros
      9

```

Figura 8: Ejecución en la consola de la línea de comando anterior.

! Debe tener presente que el paquete *qcc* no es capaz de leer dos variables, por lo cual, es por esta razón que se deben crear variables etiquetadas.

### Argumentos de la función `pareto.chart`

los argumentos a modificar en un diagrama de pareto no son muy difíciles de manejar, ya que, hay algunos argumentos que se han trabajado en gráficos anteriores. A continuación se presenta la estructura del nuevo comando junto con los argumentos a modificar paso a paso:

```

pareto.chart(Defectos,ylab='Número de defectos',
main='Diagrama de pareto para efectos en piezas',
col=colorRampPalette(c('blue','red'))(6),
plot=TRUE,las=2,ylab2='Porcentaje acumulado',
cumperc=seq(0,100,by=10))

```

- ✓ **x:** Vector con los valores especificados. En el gráfico en cuestión recibirá el nombre asignado a las variables etiquetadas que en este caso será "Defectos".
- ✓ **xlab:** Indica la etiqueta del eje X. Para este caso no será utilizado.
- ✓ **ylab:** Indica la etiqueta para el eje Y esta será declarada de la forma `ylab='Número de defectos'` para este caso.
- ✓ **main:** Nombre principal del gráfico el cual será `main='Diagrama de pareto para defectos en piezas'` en este punto.
- ✓ **col:** Los colores esta vez serán dados por un degradado entre azul y rojo. Para esto deberá ejecutar la línea `col=colorRampPalette(c('Blue','Red'))(6)`. En el vector puede indicar los colores que desee (pueden ser más de dos colores) y en el valor final la proporción de degradado, es decir, entre más grande sea el valor que coloque, más degradado se verán los colores.

- ✓ **plot:** Argumento lógico que indica a *R* si el gráfico debe generarse (declárelo como `plot=TRUE`). A diferencia de los gráficos vistos en prácticas anteriores, en la generación de diagramas de pareto en *Rstudio* en la consola se generará una tabla con las frecuencias y porcentajes acumulados y relativos para los datos correspondientes, **independientemente de que este argumento sea declarado como verdadero o falso:**

```
Pareto chart analysis for Defectos
      Frequency  Cum.Freq.  Percentage  Cum.Percent.
R. superficiales 119.000000 119.000000  62.631579   62.631579
Ruptura          37.000000 156.000000  19.473684   82.105263
Incompletas     13.000000 169.000000   6.842105   88.947368
F. inapropiadas 12.000000 181.000000   6.315789   95.263158
Otros            9.000000 190.000000   4.736842  100.000000
```

Figura 9: Tabla de frecuencias y porcentajes generados en la consola

- ✓ **las:** Argumento numérico el cual cambia la orientación de los enunciados para los todos los ejes. En este caso tendrá un valor de 2 de la forma `las=2`

Los argumentos que aparecen nuevos en este gráfico son;

- ✓ **ylab2:** Permite colocar nombre al eje vertical de la parte derecha. En este caso será `ylab2='Porcentaje acumulado'`
- ✓ **cumperc:** Es un vector que indica la escala porcentual del eje anteriormente mencionado. Para esto deberá declarar una escala de 0 a 100% con pasos de a 10 con la línea `cumperc=seq(0,100,by=10)`

! El argumento **data** no es citado en este espacio debido a que, más que un argumento, lo que indica es la forma en la cual deben introducirse los datos.

De esta manera el resultado final se puede apreciar en la *figura 10*. Como se puede evidenciar, la complejidad de este gráfico no se encuentra en los argumentos *per se* si no en la correcta utilización de los datos.

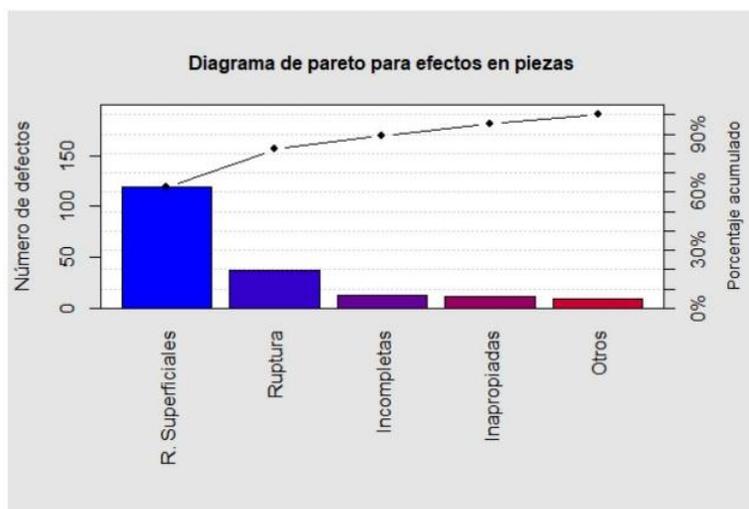


Figura 10: Diagrama de pareto resultante

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

1. Si bien el diagrama de Pareto generado da a indicar que los defectos más importantes son los rasguños superficiales, aún hace falta un análisis mucho más intensivo. Utilice la hoja de registro de datos de la *Tabla 1* para desarrollar los siguientes diagramas de Pareto que cuenten con los parámetros descritos a continuación:
  - ✓ Defectos por día turno AM
  - ✓ Defectos por día turno PM
  - ✓ Defectos por máquina turno AM
  - ✓ Defectos por máquina turno PM
2. De acuerdo al análisis realizado a los cuatro diagramas elaborados ¿Cuál cree usted que es o pueden ser las causas que están generando la mayor parte de defectos en el turno AM? ¿Esa o estas causas son las mismas que genera los defectos del turno PM? Justifique detalladamente.
3. ¿Existirá algún factor en un día en específico que haga aumentar los defectos?

## 6. Referencias

- [1] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010

# PRACTICA 5: DIAGRAMA ISHIKAWA

## 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente diagramas de causa y efecto en *R*. mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Conocimiento de los parámetros modificables y la importancia de éstos.
- ✓ Interpretar la información del gráfico resultante.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

### 3.1 Definición.

El diagrama de Ishiwaka, o también conocido como **diagrama de causa y efecto** o **diagrama de espina de pescado**, es una representación gráfica que permite poner de manifiesto las diferentes causas que influyen en un problema. Recibe su nombre en honor al experto en control de calidad *Kaoru Ishikawa*.

Un diagrama de causa y efecto está compuesto de un eje horizontal en el cual, en un extremo (generalmente el derecho) se encuentra enunciado un problema que es catalogado como el **efecto**. Es a partir de allí donde se desarrollan ramificaciones a este problema que se denominan **causas**. Los diagrama de Ishikawa son una herramienta que permite al grupo de trabajo analizar de manera conjunta y sinérgica todos los contextos que influyen en el factor problema, es por esta razón que se recomienda acudir a una lluvia de ideas para el desarrollo satisfactorio de este gráfico.

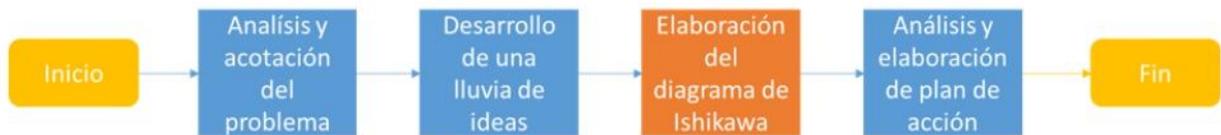


Figura 1: Diagrama de flujo a seguir en un grupo de trabajo al momento de elaborar y analizar un diagrama de causa y efecto.

## 4. Montaje y procedimiento

### Planteamiento del problema.

Para el desarrollo de esta practica, se solicita elaborar un diagrama Ishikawa en el cual se evaluen los obstáculos que le impiden a una persona bajar de peso [1]. Esta información se ha clasificado en cinco categorías y se encuentra detallada en la tabla que se evidencia a continuación:

Tabla 1: Causas y sub-causas que influyen para que una persona baje de peso.

Causas	Sub-Causas
Nutricionales	Excesos de calorías
	Exceso de porciones
	Exceso de azúcares
Genéticos	Obesidad
	Metabolismo
	Retención de líquidos
	Tiroides
Psicológicos	Ansiedad
	Compulsividad
	Depresión
Hábitos	Exceso de horas de sueño
	Mala alimentación
	Sedentarismo
Sociales	Exceso de eventos sociales
	Costumbres familiares

**Docente:** Indique al estudiante que abra el software *R studio* y que elabore un diagrama de causa y efecto (Ishikawa) con los datos de la *Tabla 1*. luego, dé la instrucción de ir realizando las modificaciones una por una a los argumentos pertinentes.

**Estudiante:** Elabore el diagrama Ishikawa de acuerdo a los datos de la tabla correspondiente. Para realizar esto deberá acudir nuevamente al paquete *Quality Control Charts* o *qcc*.

### Elaboración del código.

! Recuerde que si no se encuentra instalado el paquete *qcc* en el terminal, podrá hacerlo ejecutando la línea `install.packages('qcc')` en el área de edición de scripts, o bien, buscándolo en el entorno de utilidades haciendo click en el botón "Packages" y enseguida en "Install".

Una vez se cerciore que el paquete está instalado, recuerde llamarlo a escena mediante la forma `library('qcc')`. Cuando toda la antesala esté preparada, el comando que debe utilizar para la creación de un diagrama de Ishiwaka corresponde a la función `cause.and.effect`. Su forma general conrresponde a la siguiente estructura:

```
cause.and.effect(cause, effect, title ,cex, font)
```

A continuación se muestra en la *Figura 2* el **primer código a elaborar** en orden con sus causas efectos respectivos. Por otra lado, debajo de esta imagen se dictaminan los conceptos de los argumentos correspondientes

en esta función:

```
cause.and.effect(cause=list(Nutricionales=c("Exceso de calorías", "Exceso de porciones", "Exceso de azúcares"),
  Genéticos=c("Obesidad", "Metabolismo", "Retención de Líquidos", "Tiroides"),
  Psicológicos=c("Ansiedad", "Compulsividad", "Depresión"),
  Hábitos=c("Sedentarismo", "Mala alimentación", "Exceso en horas de sueño"),
  Sociales=c("Exceso de eventos sociales", "Costumbres familiares")),
  effect="Problemas para bajar de peso")
```

Figura 2: Puesta en marcha de los argumentos *Cause* y *effect* en el área de edición de scripts.

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

### Argumentos de la función `cause.and.effect`.

Los argumentos principales para su elaboración son:

- ✓ **cause:** En este argumento se colocarán todas las causas posibles que pueden influir en los efectos en cuestión. Para hacer esto se debe utilizar la función `list()` en la cual se deberán crear varios vectores y asignarles su respectivo nombre. En la *Figura 2* se evidencian las diferentes causas que influyen en el problema planteado y la forma correcta de su ingreso
- ✓ **effect:** Argumento en el cual se deberá escribir entre comillas el efecto a analizar. En la *Figura 2* se evidencia su forma de uso que en este caso debe citar como "Problemas para bajar de peso".

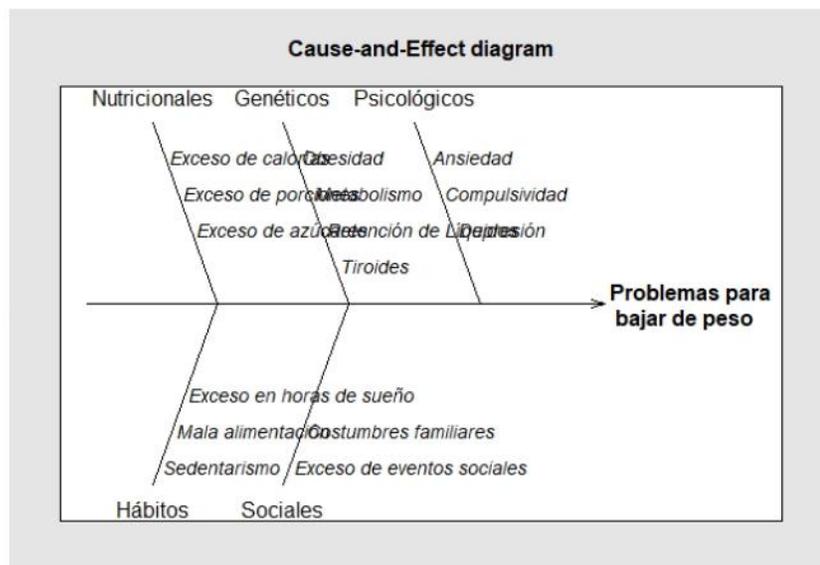


Figura 3: Diagrama de causa y efecto con valores por defecto.

Una vez se han ingresado todas las causas y los efectos, al ejecutar el comando este debe tener la siguiente

apariciencia que se muestra en la *Figura 3*. Cómo podrá observar el gráfico generado por defecto no cuenta con la mejor visibilidad. Las palabras se entrecruzan y su entendimiento es difuso.

Para poder generar una mejor apariencia en el gráfico en cuestión, se muestran los nuevos parámetros **para añadir al código actual**. Recuerde que para hacer esto deberá ir al finar del código, borrar el paréntesis y añadir una coma, y enseguida escribir esta línea de código.

```
title="Diagrama Ishikawa de factores en contra para disminuir de peso ", cex=c(1,0.7,1),
font=c(15,16,17))
```

De tal manera que la estructura del código debe estar hasta este momento como se evidencia en la *Figura 4*:

```
cause.and.effect(cause=list(Nutricionales=c("Exceso de calorías", "Exceso de porciones", "Exceso de azúcares"),
  Genéticos=c("Obesidad", "Metabolismo", "Retención de Líquidos", "Tiroides"),
  Psicológicos=c("Ansiedad", "Compulsividad", "Depresión"),
  Hábitos=c("Sedentarismo", "Mala alimentación", "Exceso en horas de sueño"),
  Sociales=c("Exceso de eventos sociales", "Costumbres familiares")),
  effect="Problemas para bajar de peso ",
  title="Diagrama Ishikawa de factores en contra para disminuir de peso ",
  cex=c(1,0.7,1),font=c(15,16,17))
```

Figura 4: Diagrama de causa y efecto con la nueva línea de código señalada en color rojo.

las modificaciones correspondientes a los argumentos serán explicados a continuación:

- ✓ **title:** Coloca título al diagrama. En este caso escriba la línea `title=' 'Diagrama Ishikawa de factores en contra para disminuir de peso'`
- ✓ **cex:** Cambia el tamaño de la letra. Si se utiliza como un vector se puede cambiar respectivamente el tamaño de los títulos de las causas, el contenido de las espinas y del efecto. Para este caso colocará `cex=c(1,0.7,1)`. Estos valores indican un aumento de forma porcentual en base al tamaño por defecto, es decir, el primer y último dígito que en este caso es **1**, indica que se mantendrá el tamaño base, por otra parte, la cifra **0.7** indica que el tamaño de la letra en las espinas tuvo una reducción del 30%.
- ✓ **font:** Cambia la tipografía al diagrama. Si se utiliza como un vector cambia la tipografía en el mismo orden que en el argumento anterior (título de causas, contenido de espinas y efecto). Este parámetro cuenta con las tipografías básicas de **Arial MT**, **Perpetua**, **Typewriter** y **Century Gothic**. Sin embargo, se le puede indicar a *Rstudio* el uso de letras griegas y símbolos. Los valores que se pueden utilizar y sus respectivos atributos se encuentran especificados en la *Tabla 2*. Para el diagrama que se está realizando, se declarará `font=c(15,16,17)` respectivamente.

Tabla 2: Parámetros de tipografía para el argumento *font*

Dígito	Tipografía	Característica
1, 18 y 20 en adelante	Arial MT	Sin negrita ni cursiva
2		Negrita
3		Cursiva
4		Negrita y cursiva
5	Alfabeto griego	Sin negrita ni cursiva
6	Perpetua	Sin negrita ni cursiva
7		Negrita
8		Cursiva
9		Negrita y cursiva
10	Typewriter	Sin negrita ni cursiva
11		Negrita
12		Cursiva
13		Negrita y cursiva
14	Century Gothic	Sin negrita ni cursiva
15		Negrita
16		Cursiva
17		Negrita y cursiva
19	Symbol	Sin negrita ni cursiva

Ya para finalizar, se puede agregar espacios verticales para dividir el texto añadiendo una contrabarra o barra invertida con la letra **n** en el lugar que se desea dividir. En la siguiente figura se evidencia el uso de esta herramienta:

```
cause.and.effect(cause=list(Nutricionales=c("Exceso \n de calorías", "Exceso de \n porciones", "Exceso de \n zúcares"),
  Genéticos=c("Obesidad", "Metabolismo", "Retención de \n Líquidos", "Tiroides"),
  Psicológicos=c("Ansiedad", "Compulsividad", "Depresión"),
  Hábitos=c("Sedentarismo", "Mal \n alimentación", "Exceso en horas \n de sueño"),
  Sociales=c("Exceso de eventos \n sociales", "Costumbres familiares")),
  effect="Problemas para \n bajar de peso",title="Diagrama Ishikawa de factores en contra para disminuir de pesc
  cex=c(1,0.7,1),font=c(15,16,17))
```

Figura 5: Uso de la contrabarra y letra n para crear espacios verticales en el diagrama de ishiwaka.

De esta manera, una vez se han realizado las respectivas modificaciones a los argumentos restantes y se han arreglado los problemas de las letras cruzadas con las cotrabarras y **n** el diagrama de ishiwaka realizado tiene el siguiente aspecto:

! Si está trabajando en *Rstudio Cloud* a pesar de realizar estas modificaciones, aún se verán las palabras sobrepuestas, con lo que para ver correctamente el gráfico deberá hacer click en el botón "Zoom".

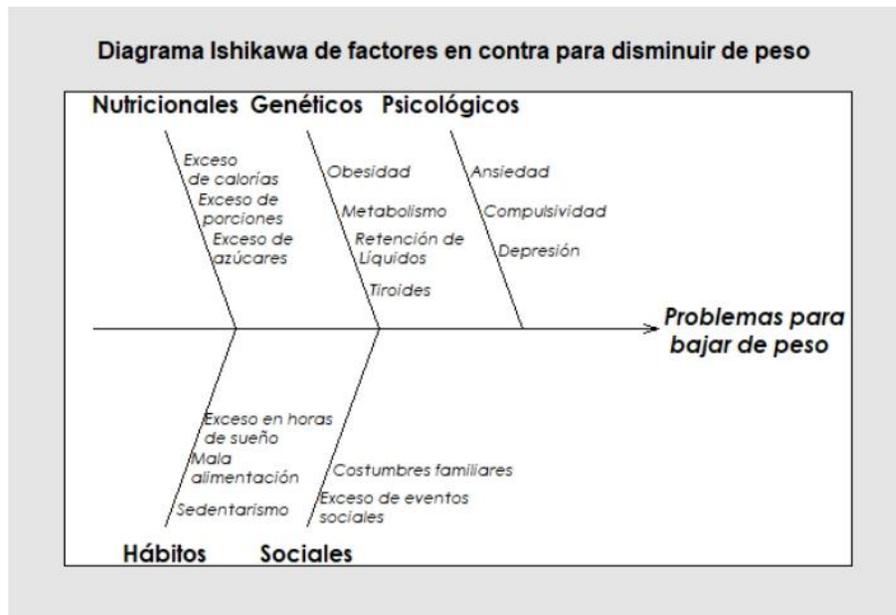


Figura 6: Diagrama de causa y efecto (ishiwaka) Final.

Como se puede observar en la *Figura 6*, el diagrama ha tenido una mejora significativa con respecto al primer gráfico.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuesta deben estar relacionadas con lo observado durante la práctica.

1. De acuerdo al diagrama de ishiwaka elaborado ¿Cuál cree usted que es el factor que más genera impacto en el problema principal?. Justifique su respuesta.
2. Desarrolle un plan de acción en el cual pueda evidenciarse la mitigación para cada una de las causas (nutricionales, genéticos, psicológicos, hábitos y sociales). Luego indique los beneficios que estas oportunidades de mejora tiene en corto, mediano y largo plazo (1 mes, 6 meses y un año respectivamente).
3. Realice una lluvia de ideas con sus compañeros y luego elabore un diagrama de causa y efecto para el siguiente caso:
  - ✓ ¿Cuáles son las causas por las que una persona llega tarde a una cita?.
4. Una vez realizado el diagrama de flujo, realice el mismo procedimiento aplicado en el punto 2 para este caso.

## 6. Referencias

- [1] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010

## PRACTICA 6: CARTA DE CONTROL X - R

### 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control para variables  $X - R$  mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Conocimiento de elementos básicos para la elaboración e interpretación de los datos en el software *Rstudio* correspondiente a cartas de control.
- ✓ Interpretación y análisis de la información del gráfico resultante.

### 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

### 3. Teoría relacionada

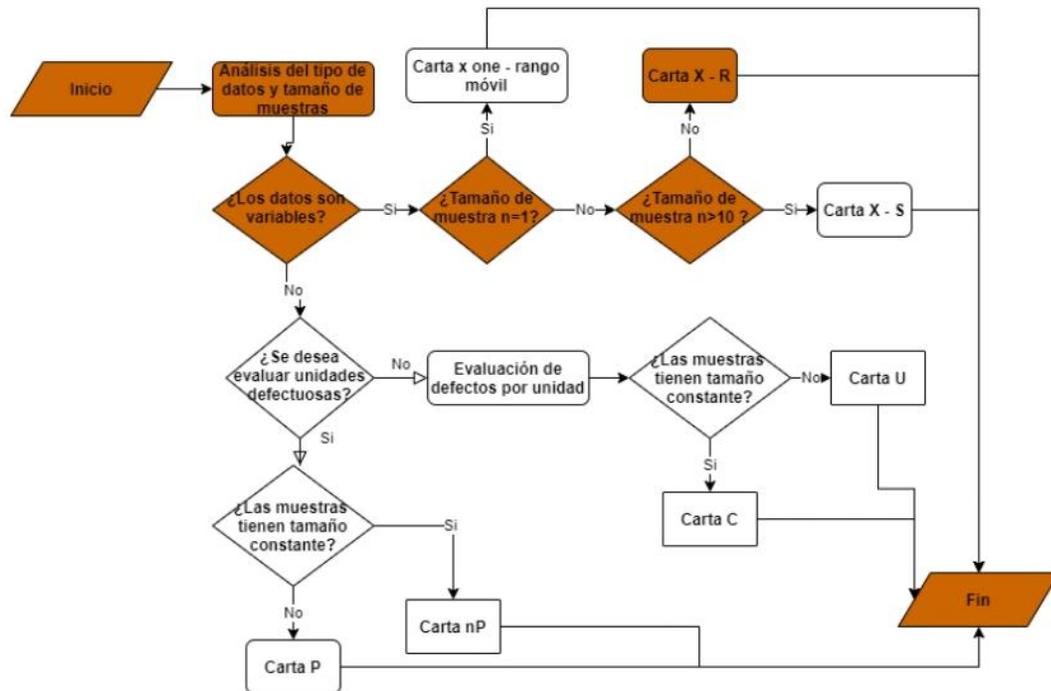


Figura 1: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta  $\bar{X} - R$ .

### 3.1 Definición.

Las cartas de control son una herramienta fundamental para el control estadístico de la calidad las cuales tienen la gran bondad de dar monitoreo y control a las especificaciones de un proceso. Estas se clasifican en dos tipos importantes:

- ✓ **Cartas de control para variables:** Permiten la medición de manera cuantitativa de una magnitud en concreto (temperatura, longitud, viscosidad, entre otros.) La idea es realizar un control al cumplimiento de los requerimientos en base a una métrica especificada.
- ✓ **Cartas de control por atributos:** Estas permiten dictaminar si el proceso está bajo control estadístico en base a la conformidad y no conformidad de este en función de ciertos parámetros cualitativos.

El concepto de la carta de control fue desarrollado en 1920 por el físico, ingeniero y estadístico estadounidense *Walter A. Shewhart* de Bell Telephone Laboratories.[1]

#### Ventajas de las cartas de control [1]

1. Son una herramienta magistral en pro de la productividad ya que dimite en el hecho de reducir los desperdicios.
2. Previenen defectos.
3. Evitan el ajuste innecesario del proceso.
4. Son herramientas precisas ya que identifican con claridad la causas fortuitas y asignables. Esto es fundamental ya que, un operario no tiene la sensibilidad para reconocer esta clase de ruido blanco.
5. Brinda información relacionada con la capacidad del proceso.

#### Cartas de control X - R

La cartas de control para medias  $\bar{X}$  y rangos  $R$  trabajan de forma armoniosa con el fin de monitorear procesos en los que se utilicen tamaños de muestras  $n$  menores a 10. La carta  $\bar{X}$  permite el monitoreo de la media de las muestras en el proceso. Por otra parte, la carta  $R$  permite observar la variabilidad de este. Es importante realizar la elaboración de estos gráficos de control en conjunto debido a que hay procesos en los que la media de las muestras se encuentra bajo control, sin embargo su variabilidad es inestable, y otros en los que la variabilidad está controlada y el promedio de las muestras puede estar siendo afectado. Esto se puede apreciar más claramente en la *Figura 2*.

#### Paso a paso para elaborar cartas X - R.

El monitoreo del control estadístico de calidad a través de las cartas de control debe seguir los siguientes pasos:

1. Calcular los límites de control. Las ecuaciones correspondiente para tales fines son:

**Limites de control para gráfico X      Limites de control para gráfico R**

$$UCL = \bar{\bar{X}} + A_2\bar{R}$$

$$CL = \bar{\bar{X}}$$

$$LCL = \bar{\bar{X}} - A_2\bar{R}$$

$$UCL = D_4\bar{R}$$

$$CL = \bar{R}$$

$$LCL = D_3\bar{R}$$

! Los valores de las constantes  $A_2$ ,  $D_3$  y  $D_4$  los puede encontrar en la *Tabla de valores para la construcción de cartas de control* disponible en libros de control estadístico de calidad o bien en internet. Tenga presente el correspondiente valor para el tamaño de la muestra especificado.

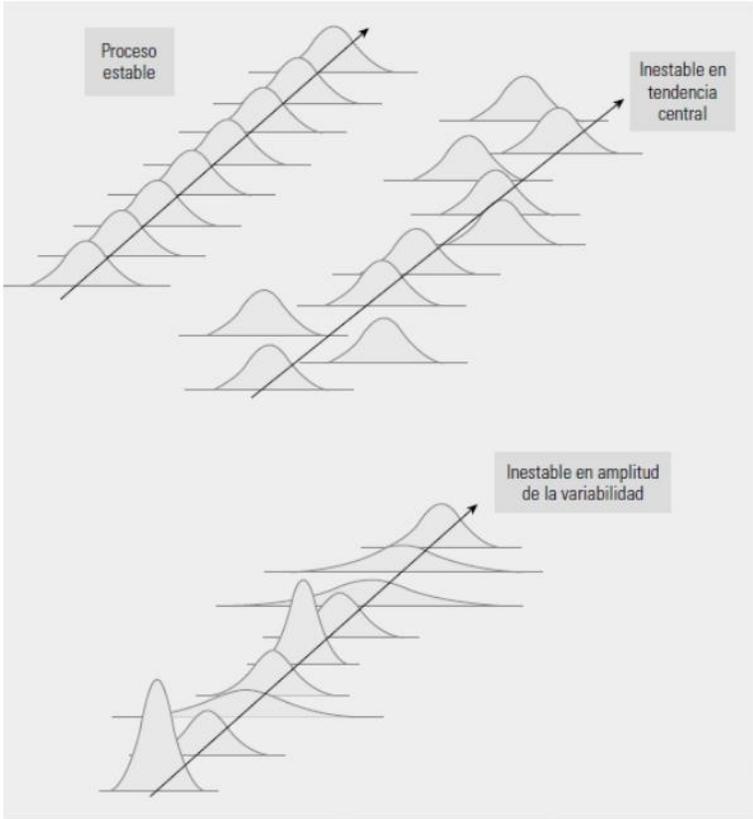


Figura 2: Representación gráfica de la variabilidad en la media y rango de acuerdo a una distribución normal.[2]

- 2. Realizar un análisis a los puntos de las cartas de control, verificando las causas fortuitas y revisando detalladamente las causas asignables del proceso. Si se encuentran las causas asignables que están ocasionando inconvenientes en el proceso estas deben ser eliminadas.

! El siguiente punto no deberá ejecutarlo en *Rstudio*. Sólomente se explica el caso en el que llegue a requerir eliminar puntos para calcular nuevamente los límites de control.

- 3. Una vez se han eliminado las causas asignables, se deben calcular nuevamente los límites de control y enseguida, elaborar las respectivas cartas. En *Rstudio* para poder realizar este paso deberá realizar una modificación al conjunto de datos con el que ha trabajado, esto es, elaborar un nuevo dataframe con los datos extraídos. Esto se puede realizar haciendo uso de la línea de comando:

```
b <- a[-c(n1, n2),]
```

Donde **b** es el nombre asignado al dataframe con los datos extraídos y **a** es el conjunto de datos originales. Para extraer los datos deberá escribir un vector que contenga la posición de los datos que desea

eliminar entre llaves precedido del signo menos ( $-$ ) y separado de una coma. Esta estructura dentro de las llaves lleva la forma [filas, columnas] con lo que se está indicando eliminar esas filas en cuestión, pero ninguna columna. Es por esta razón que en la estructura evidenciada anteriormente no se coloca nada después de la coma. Ahora bien, si desea eliminar un segmento de datos puntuales puede hacer uso de dos puntos ( $:$ ) en donde deberá colocar el dígito de las posiciones de inicio y fin del intervalo  $n_1$  y  $n_2$  respectivamente. De esta forma el comando a usar queda con el siguiente aspecto:

$$b < -a[-c(n_1:n_2),]$$

Una vez realizado este procedimiento se procede a repetir el ploteo de los gráficos con el nuevo dataframe haciendo uso de la función `qcc` la cual será explicada a continuación.

#### 4. Montaje y procedimiento

**Docente:** Indique al estudiante que abra el software *R studio* y que elabore una carta de control de medias  $\bar{X}$  y una de rangos  $R$  para los datos la *Tabla 1*, realizando la correspondiente contextualización del problema planteado y la explicación paso a paso del código correspondiente.

**Estudiante:** Elabore las cartas de control  $\bar{X}$  y  $R$  de acuerdo a los datos de la tabla correspondiente. Para la elaboración de estos gráficos de control se trabajará en el siguiente caso:

*Tabla 1: Datos correspondientes a las 25 muestras para anillos de pistones. [1]*

Número de muestra	obs1	obs2	obs3	obs4	obs5	Media subgrupo	Rango subgrupo
1	74.030	74.002	74.019	73.992	74.008	74.010	0.038
2	73.995	73.992	74.001	74.011	74.004	74.001	0.019
3	73.988	74.024	74.021	74.005	74.002	74.008	0.036
4	74.002	73.996	73.993	74.015	74.009	74.003	0.022
5	73.992	74.007	74.015	73.989	74.014	74.003	0.026
6	74.009	73.994	73.997	73.985	73.993	73.996	0.024
7	73.995	74.006	73.994	74.000	74.005	74.000	0.012
8	73.985	74.003	73.993	74.015	73.988	73.997	0.030
9	74.008	73.995	74.009	74.005	74.004	74.004	0.014
10	73.998	74.000	73.990	74.007	73.995	73.998	0.017
11	73.994	73.998	73.994	73.995	73.990	73.994	0.008
12	74.004	74.000	74.007	74.000	73.996	74.001	0.011
13	73.983	74.002	73.998	73.997	74.012	73.998	0.029
14	74.006	73.967	73.994	74.000	73.984	73.990	0.039
15	74.012	74.014	73.998	73.999	74.007	74.006	0.016
16	74.000	73.984	74.005	73.998	73.996	73.997	0.021
17	73.994	74.012	73.986	74.005	74.007	74.001	0.026
18	74.006	74.010	74.018	74.003	74.000	74.007	0.018
19	73.984	74.002	74.003	74.005	73.997	73.998	0.021
20	74.000	74.010	74.013	74.020	74.003	74.009	0.020
21	73.982	74.001	74.015	74.005	73.996	74.000	0.033
22	74.004	73.999	73.990	74.006	74.009	74.002	0.019
23	74.010	73.989	73.990	74.009	74.014	74.002	0.025
24	74.015	74.008	73.993	74.000	74.010	74.005	0.022
25	73.982	73.984	73.995	74.017	74.013	73.998	0.035

Los anillos para pistones de un motor de automóvil se producen mediante un proceso de fundición. Quiere establecerse el control estadístico del diámetro interior de los anillos fabricados con este proceso utilizando las cartas  $\bar{X}$  y  $R$ . se toman 25 muestras, cada una de tamaño cinco, cuando se considera que el proceso está bajo control.

### Importación de archivos e instalación del paquete `qcc`.

Para la elaboración de las cartas de control, primero deberá importar la *Tabla 1* a *Rstudio* como un archivo *TSV* o **Texto delimitado por tabulaciones**. Recuerde que para hacer esto debe ir a *Excel* y elaborarla allí. luego debe dar click en el botón “guardar como” en la pestaña *Archivo* y una vez está abierto el cuadro de texto en la parte inferior de este en la parte de “tipo” podrá seleccionar “texto (delimitado por tabulaciones)”.

! Si está trabajando en *Rstudio Cloud* suba el archivo directamente como un archivo de *Excel*.

! Recuerde que si no se encuentra instalado el paquete `qcc` en el terminal, podrá hacerlo ejecutando la línea `install.packages('qcc')` en el área de edición de scripts, o bien, buscándolo en el entorno de utilidades haciendo click en el botón “Packages” y enseguida en “Install”.

### Elaboración del código.

Enseguida, deberá hacer la puesta en escena del paquete `qcc` el cual tiene la gran virtud de ofrecer una variedad de gráficos de control. En cuanto al código a ejecutar, deberá seguir en orden la siguiente línea de comandos que se explican a continuación:

```
attach(Datos)
names(Datos)
Ejercicio6<-data.frame(obs1,obs2,obs3,obs4,obs5)
Ejercicio6
```

```
library('qcc')
```

```
qcc(Ejercicio6,type = 'xbar')
qcc(Ejercicio6,type = 'R')
```

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escribálas en *Rstudio*.

1. Una vez ha importado los datos, utilice la función `attach` para llamar el conjunto de datos y no tener que requerir para más adelante de llamar alguna columna con el signo pesos.
2. Haga uso de la función `names` para que pueda tener conocimiento del nombre respectivo para cada una de la columnas.
3. Deberá crear un nuevo dataframe debido a que para poder realizar las cartas de control, es menester eliminar la columna *Número de muestra* que pudo apreciar en el inciso anterior con lo que deberá decla-

rar de la forma `Ejercicio6<-data.frame(obs1,obs2,obs3,obs4,obs5)` para este caso. Esto debido a que si se realiza el código para plotear el gráfico con esta columna, las cartas de control serán una sucesión de puntos ascendente, y no la muestra de los puntos deseados.

4. Ejecute el nombre asignado al dataframe para comprobar que ha sido eliminada la columna de *Número de muestra* (En este caso nombrado como “Ejercicio 6”). Con esto ya los datos tendrán el tratamiento inicial para la elaboración de los gráficos.
5. Partiendo de la base de que la librería *qcc* ya está instalada en el terminal, haga la puesta en escena de esta haciendo uso de la función `library()` y escribiendo el nombre de la librería entre paréntesis.
6. La función para realizar los gráficos de control se llama de forma simple como `qcc`. La forma general de esta función en donde se evidencian los argumentos más usados es:

! Cuando se habla de la forma general de una función, esto sólo indica la forma correcta de su uso y el tipo de argumentos que esta utiliza (lógico, numérico, vector, entre otros). **No se debe ejecutar esta línea.**

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

En su forma más básica de uso consta de los dos siguientes argumentos:

- ✓ **data:** Conjunto de datos correspondiente para la elaboración de cartas de control.
- ✓ **type:** Son los tipos de cartas de control permitidas en el paquete *qcc*. En el caso de las cartas de control para variables se pueden realizar las que se mencionan a continuación:
  - **xbar** : Corresponde a la carta de control de medias
  - **R**: Carta de control de rangos.
  - **S**: Carta de control de desviación estándar.
  - **xbar.one**: Carta de control para mediciones individuales.

En ese orden de ideas, ejecute las dos últimas líneas para poder obtener los respectivos gráficos de control.

! Este argumento también permite la elaboración de cartas de control por atributos (p, np, c, u y g) sin embargo, estas se explicarán en detalle en las siguientes prácticas.

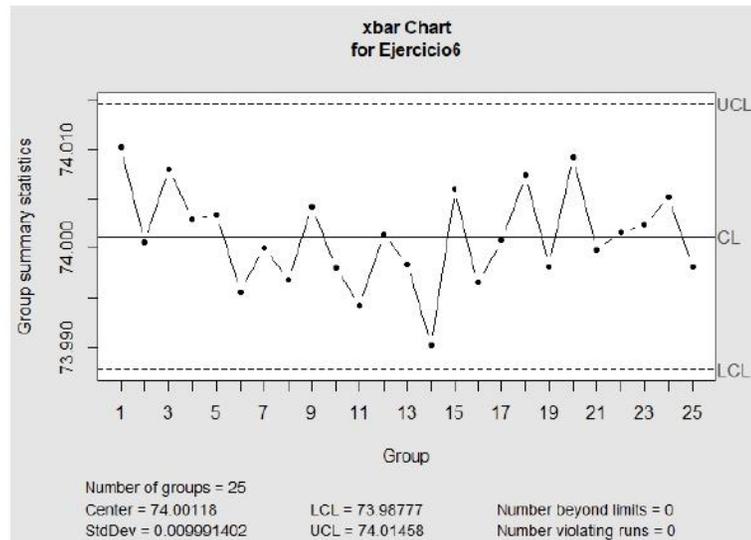


Figura 3: Carta de control de medias  $\bar{X}$ .

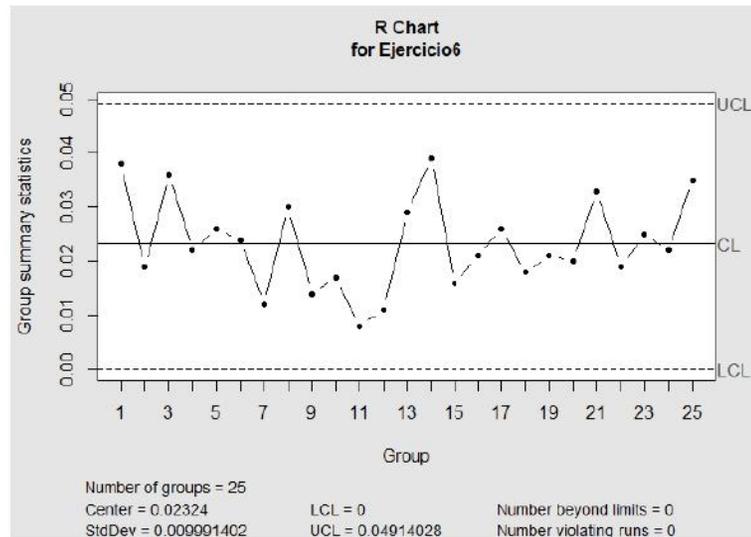


Figura 4: Carta de control de rangos  $R$ .

### Atributos estadísticos.

La figuras resultantes de los códigos respectivos deben tener la apariencia evidenciada en las *Figuras 3 y 4* respectivamente. Una vez ejecuta las líneas, además de obtener los gráficos respectivos, también logrará detallar en la consola información estadística sobre las respectivas cartas de control. Esto se puede apreciar en la *Figura 4* en la cual se especifica todos los datos que hacen parte de los gráfico de control elaborados. En total se despliega información de once (11) atributos los cuales son:

```

> qcc(Ejercicio6,type = "xbar")
List of 11
 $ call      : language qcc(data = Ejercicio6, type = "xbar")
 $ type      : chr "xbar"
 $ data.name : chr "Ejercicio6"
 $ data      : num [1:25, 1:5] 74 74 74 74 74 ...
 .. attr(*, "dimnames")=List of 2
 $ statistics: Named num [1:25] 74 74 74 74 74 ...
 .. attr(*, "names")= chr [1:25] "1" "2" "3" "4" ...
 $ sizes     : int [1:25] 5 5 5 5 5 5 5 5 5 ...
 $ center    : num 74
 $ std.dev   : num 0.00999
 $ nsigmas   : num 3
 $ limits    : num [1, 1:2] 74 74
 .. attr(*, "dimnames")=List of 2
 $ violations:List of 2
 - attr(*, "class")= chr "qcc"
> qcc(Ejercicio6,type = "R")
List of 11
 $ call      : language qcc(data = Ejercicio6, type = "R")
 $ type      : chr "R"
 $ data.name : chr "Ejercicio6"
 $ data      : num [1:25, 1:5] 74 74 74 74 74 ...
 .. attr(*, "dimnames")=List of 2
 $ statistics: Named num [1:25] 0.038 0.019 0.036 0.022 0.026 ...
 .. attr(*, "names")= chr [1:25] "1" "2" "3" "4" ...
 $ sizes     : int [1:25] 5 5 5 5 5 5 5 5 5 ...
 $ center    : num 0.0232
 $ std.dev   : num 0.00999
 $ nsigmas   : num 3
 $ limits    : num [1, 1:2] 0 0.0491
 .. attr(*, "dimnames")=List of 2
 $ violations:List of 2
 - attr(*, "class")= chr "qcc"

```

 Datos para carta de medias X  
 Datos para carta de rangos R

Figura 5: Datos estadísticos para las cartas de control  $\bar{X}$  y  $R$  generados en la consola.

- ✓ **call:** Muestra el código el cual llama al gráfico.
- ✓ **type:** Especifica el tipo de gráfico que se está utilizando.
- ✓ **data.name:** Esta opción muestra en la consola el nombre del conjunto de datos.
- ✓ **statistics:** Realiza los cálculos de acuerdo al tipo de carta especificado para todos los subgrupos. Por ejemplo, en el caso que se estén observando estos atributos para la carta  $R$  se puede apreciar en la *Figura 4* que en la consola se encuentran los rangos correspondientes para cada una de las 25 muestras. Si por otro lado, se utiliza la carta de control para medias, en este atributo se calcularán las medias para cada una de las 25 muestras.
- ✓ **sizes:** Se aprecia el tamaño de las muestras para cada uno de los subgrupos.
- ✓ **center:** Calcula el valor de la línea central para la carta de control respectiva.
- ✓ **std.dev:** Indica el valor de la desviación estándar especificado.
- ✓ **nsigmas:** Muestra el número de sigmas establecido para los límites de control. Tenga presente que este valor es 3 por defecto.
- ✓ **limits:** Muestra los valores de los límites de control establecidos.



```
media<-qcc(Ejercicio6,type = 'xbar',xlab='Subgrupos',ylab='',
title='Carta de control X de anillos para pistones',
plot=TRUE,axes.las=1)
```

```
rango<-qcc(Ejercicio6,type = 'R', xlab='Subgrupos',ylab='',
title='Carta de control R de anillos para pistones',
plot=TRUE,axes.las=1)
```

- ✓ **xlab:** Indica la etiqueta del eje X. En el caso de este argumento deberá declarar **para ambas cartas de control** de la forma `xlab=subgrupos`
- ✓ **ylab:** Indica la etiqueta para el eje Y. Si declara el argumento de la forma `ylab = ''` no saldrá nada en el gráfico.
- ✓ **ylim:** Vector numérico que especifica los límites del eje Y. No es necesario declarar este argumento ya que por defecto la carta de control ya dictamina los límites de control.
- ✓ **title:** Argumento que permite colocar el título a las cartas de control. Esta vez especifique para los gráficos respectivamente como `title='Carta de control X de anillos para pistones'` y `title='Carta de control R de anillos para pistones'`.
- ✓ **plot:** Argumento lógico que permite la elaboración del gráfico al ser declarado como verdadero. En ambas cartas será declarado como `plot=TRUE`
- ✓ **axes.las:** Argumento numérico el cual cambia la orientación de los valores de los ejes. Recuerde que puede hacer uso solamente los valores de 0, 1, 2 y 3. Dictamine de la forma `las=1` para ambos gráficos de control.

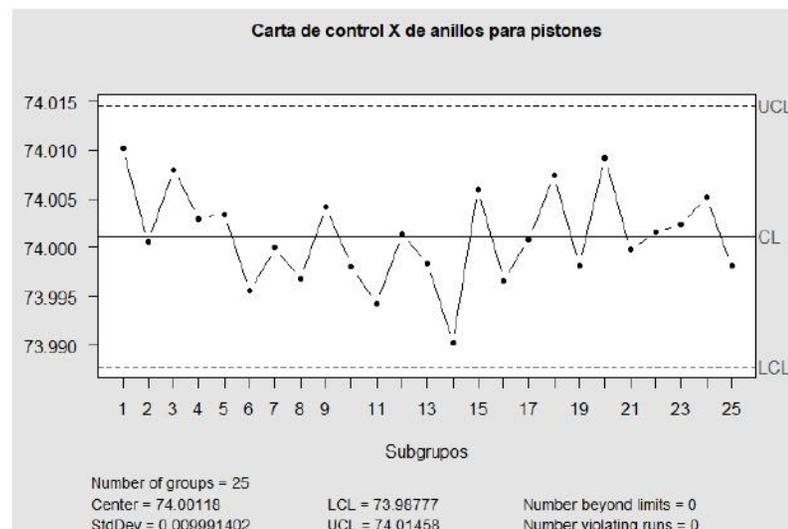


Figura 7: Carta de control  $\bar{X}$  con las modificaciones a argumentos iniciales.

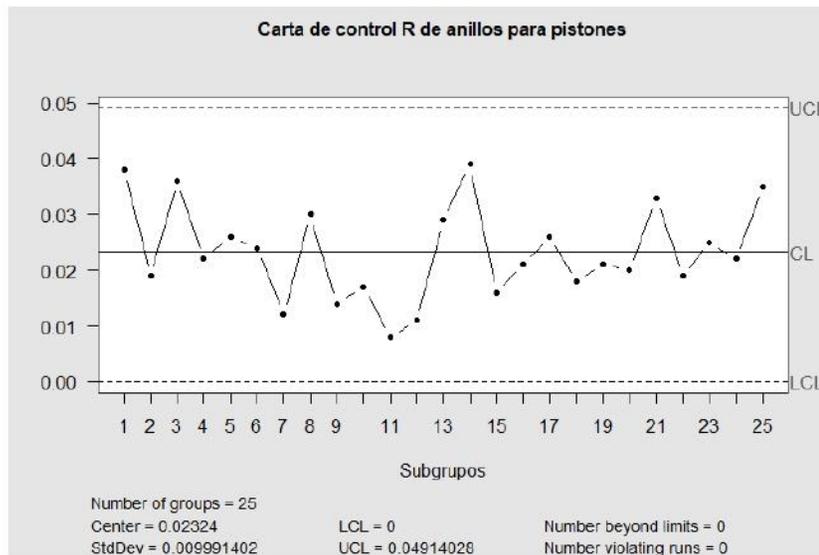


Figura 8: Carta de control de rangos  $R$ . con modificaciones a argumentos iniciales.

De esta manera, una vez realizados estos ajustes iniciales los respectivos resultados se aprecian en las Figuras 6 y 7 respectivamente.

! La función *gcc* para la generación de cartas de control cuenta con una amplia variedad de argumentos y se apoya de algunas funciones mucho más complejas. Estas se verán a profundidad en próximas guías.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

A partir de la siguiente información, resuelva los siguientes puntos:

*En una empresa fabricante de corcholatas o tapas metálicas para bebidas gaseosas, un aspecto importante es la cantidad de PVC que lleva cada corcholata, la cual determina el espesor de la película que hace que la bebida quede bien cerrada. El peso de los gránulos de PVC debe estar entre 212 y 218 mg. Si el peso es menor a 212, entonces, entre otras cosas, la película es muy delgada y eso puede causar fugas de gas en la bebida. Si el peso es mayor que 218 g, entonces se gasta mucho PVC y aumentan los costos. Para asegurar que se cumplen con especificaciones, se usa ordinariamente una carta de control: cada 30 minutos se toma una muestra de cuatro gránulos consecutivos de PVC y se pesan. En la **Tabla 2** se muestran las últimas 25 medias y rangos obtenidas del proceso.*

1. Calcule los límites de una carta  $X - R$  y obtenga las cartas.
2. Realice la respectiva interpretación de las cartas.
3. ¿El proceso muestra una estabilidad o estado de control estadístico razonable?

Tabla 2: Valores de media y rango para la cantidad de PVC para corcholatas.

Subgrupo	Media	Rango
1	214.18	2.5
2	213.48	2.7
3	213.98	2.2
4	214.12	1.8
5	214.46	2.5
6	213.38	2.7
7	213.56	2.3
8	214.08	1.8
9	213.72	2.9
10	214.64	2.2
11	213.92	2.4
12	213.96	3.6
13	214.20	0.4
14	213.74	3.2
15	214.26	1.2
16	214.18	2.2
17	214.00	1.0
18	213.60	2.0
19	214.20	2.7
20	214.38	0.8
21	213.78	2.0
22	213.74	1.6
23	213.32	2.4
24	214.02	3.2
25	214.24	1.1

## 6. Referencias

- [1] D. V. Montgomery, Douglas C y Verbeek. *Control estadístico de la calidad*. Limusa Wiley, 2004.
- [2] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010.

## PRACTICA 7: CARTA DE CONTROL X - S

### 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control para variables  $\bar{X} - S$  mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Conocimientos pertinentes para la aplicación de las funciones *qcc.groups* y *qcc.options* en cartas de control.
- ✓ Interpretación y análisis de la información del gráfico resultante.

### 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

### 3. Teoría relacionada

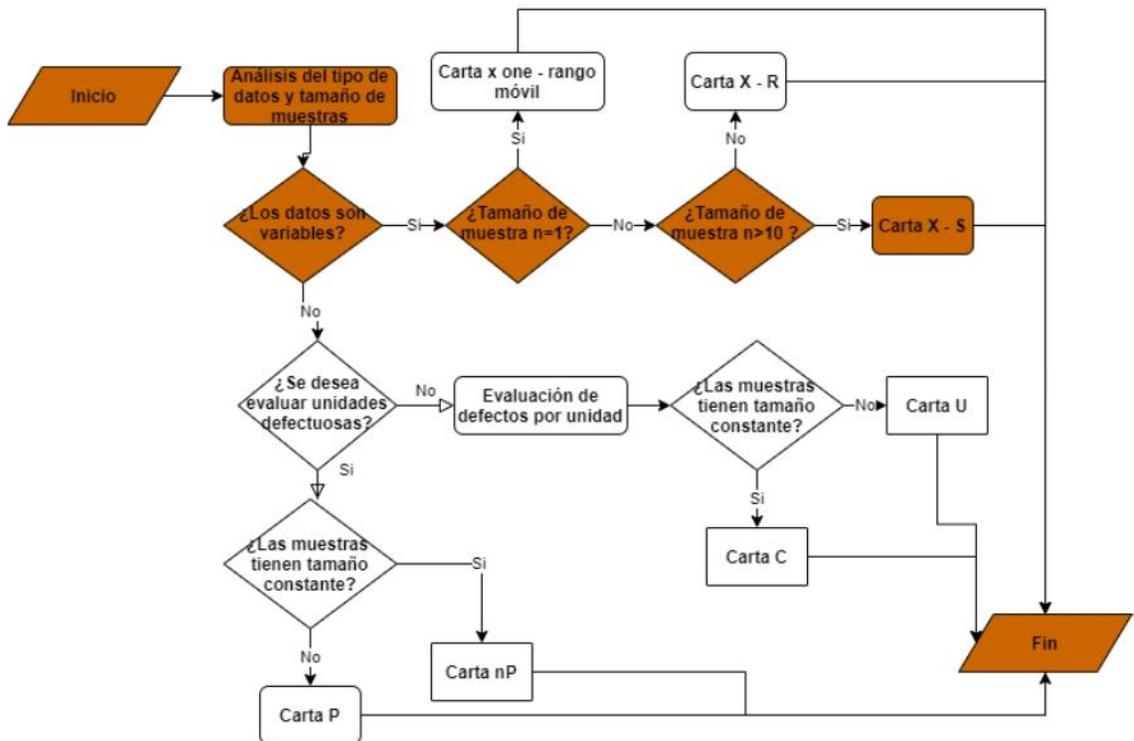


Figura 1: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta  $\bar{X} - S$ .

### 3.1 Definición.

Las cartas de control  $\bar{X}$  y  $S$  son el otro caso práctico de uso para monitoreo de procesos mediante variables. Estas cartas de control se utilizan en los siguientes casos:

1. El tamaño de las muestras  $n$  es superior a 10 ó 12.
2. Se presenta una variabilidad en el tamaño de las muestras, es decir, no todas las muestras presentan el mismo tamaño, lo cual hace que los límites de control también cambien cada que se grafica un punto en las cartas de control.[1]

Las respectivas ecuaciones para la obtención de los límites de control son:

**Límites de control para gráfico  $\bar{X}$**       **Límites de control para gráfico  $S$**

$$UCL = \bar{\bar{X}} + A_3\bar{S}$$

$$CL = \bar{\bar{X}}$$

$$LCL = \bar{\bar{X}} - A_3\bar{S}$$

$$UCL = B_4\bar{S}$$

$$CL = \bar{S}$$

$$LCL = B_3\bar{S}$$

! Los valores de las constantes  $A_3$ ,  $B_3$  y  $B_4$  los puede encontrar en la *Tabla de valores para la construcción de cartas de control* disponible en libros de control estadístico de calidad o bien en internet. Tenga presente el correspondiente valor para el tamaño de la muestra especificado.

## 4. Montaje y procedimiento

**Docente:** Indique al estudiante que abra el software *R studio* y que elabore una carta de control de medias  $\bar{X}$  y una de desviaciones  $S$  para los datos del dataframe *pistonrings* integrado en el paquete *qcc*. A la postre, vaya realizando la explicación paso a paso de la función *qcc.groups* y *qcc.options*.

**Estudiante:** Elabore las cartas de control  $\bar{X}$  y  $S$  de acuerdo a los datos. Para ello deberá ejecutar paso a paso la línea de código respectiva que se explicará paso a paso a continuación:

**Elaboración del código.**

! Recuerde que si no se encuentra instalado el paquete *qcc* en el terminal, podrá hacerlo ejecutando la línea `install.packages('qcc')` en el área de edición de scripts, o bien, buscándolo en el entorno de utilidades haciendo click en el botón "Packages" y enseguida en "Install".

```
library('qcc')
pistonrings
data('pistonrings')
attach(pistonrings)
arreglo<-qcc.groups(diameter,sample)
qcc(arreglo,type='xbar')
qcc(arreglo,type='S')
```

1. Realice la labor primordial de hacer la puesta en escena del paquete *qcc* mediante el uso de la función `library`.

- Identifique los tipos de datos que dispone el dataframe *pistonrings* simplemente escribiendo su nombre y ejecutando esa línea. Se podrá percatar que este cuenta con una columna llamada *sample*, la cual indica a qué subgrupo pertenece cada muestra. En total hay cuarenta (40) subgrupos conformados cada uno por cinco (5) muestras. Esto será importante tenerlo en cuenta a la hora de ejecutar la función *qcc.options* más adelante.
- Enseguida, haciendo uso de la función la función **data** proceda a realizar el llamado del dataframe *pistonrings*.
- Haga uso de la función **attach** para trabajar sin necesidad de utilizar los signos \$ al momento de llamar columnas de interés.
- Deberá asignar un nombre (en este caso será “arreglo”) a la función *qcc.groups*. Esta tiene la finalidad de adaptar un conjunto de datos de acuerdo a alguna característica requerida. (Agrupar temperaturas en el año de acuerdo a cada mes, clasificar un tipo de pieza de acuerdo a la máquina que la elaboró, entre otros). Esto resulta fundamental debido a que es menester realizar un tipo de arreglo en el dataframe para que la función *qcc* sea capaz de realizar el gráfico de control y los cálculos respectivos. Su forma general es:

```
qcc.groups(data, sample)
```

Y los argumentos de esta función son:

- ✓ **data**: Es el conjunto de datos.
- ✓ **sample**: Es el parámetro que indicará la agrupación correspondiente.

En este orden de ideas, lo que se estará realizando en esta función es simplemente agrupar los datos de la columna *diameter* en base a los grupos expuestos en la columna *sample* con lo que la línea de código para este paso quedará de la forma como se evidencia arriba junto a los demás comandos como `arreglo<-qcc.groups(diameter, sample)`.

```
> arreglo
  [,1] [,2] [,3] [,4] [,5]
1 74.030 74.002 74.019 73.992 74.008
2 73.995 73.992 74.001 74.011 74.004
3 73.988 74.024 74.021 74.005 74.002
4 74.002 73.996 73.993 74.015 74.009
5 73.992 74.007 74.015 73.989 74.014
6 74.009 73.994 73.997 73.985 73.993
7 73.995 74.006 73.994 74.000 74.005
8 73.985 74.003 73.993 74.015 73.988
9 74.008 73.995 74.009 74.005 74.004
10 73.998 74.000 73.990 74.007 73.995
11 73.994 73.998 73.994 73.995 73.990
12 74.004 74.000 74.007 74.000 73.996
13 73.983 74.002 73.998 73.997 74.012
14 74.006 73.967 73.994 74.000 73.984
15 74.012 74.014 73.998 73.999 74.007
16 74.000 73.984 74.005 73.998 73.996
17 73.994 74.012 73.986 74.005 74.007
18 74.006 74.010 74.018 74.003 74.000
19 73.984 74.002 74.003 74.005 73.997
20 74.000 74.010 74.013 74.020 74.003
21 73.988 74.001 74.009 74.005 73.996
22 74.004 73.999 73.990 74.006 74.009
23 74.010 73.988 73.990 74.000 74.014
```

Figura 2: Nuevo arreglo de datos mediante la función *qcc.groups*

- De esta manera al ejecutar el nombre asignado a la función `qcc.groups` se tendrá organizada la información de una forma más cómoda para que *Rstudio* realice las respectivas cartas de control ejecutando las líneas que se explicarán en el siguiente punto. Puede visualizar el resultado del comando `arreglo<-qcc.groups(diameter, sample)` en la consola ejecutando esta línea o bien yendo al entorno de ambientes en la parte superior izquierda de la pantalla y haciendo click sobre el nombre. Con lo cual se abrirá una nueva ventana en el área de edición de scripts.
- Finalmente, realice los gráficos de control respectivos haciendo uso de la función `qcc`. Para ambas cartas utilice en el argumento `data` el dataset nombrado como “arreglo”; luego en `type` escriba “xbar” y “S” respectivamente de tal forma que las líneas tengan la estructura `qcc(arreglo,type='xbar')` y `qcc(arreglo,type='S')`. Con esto los resultados se evidencian en la *Figura 3*:

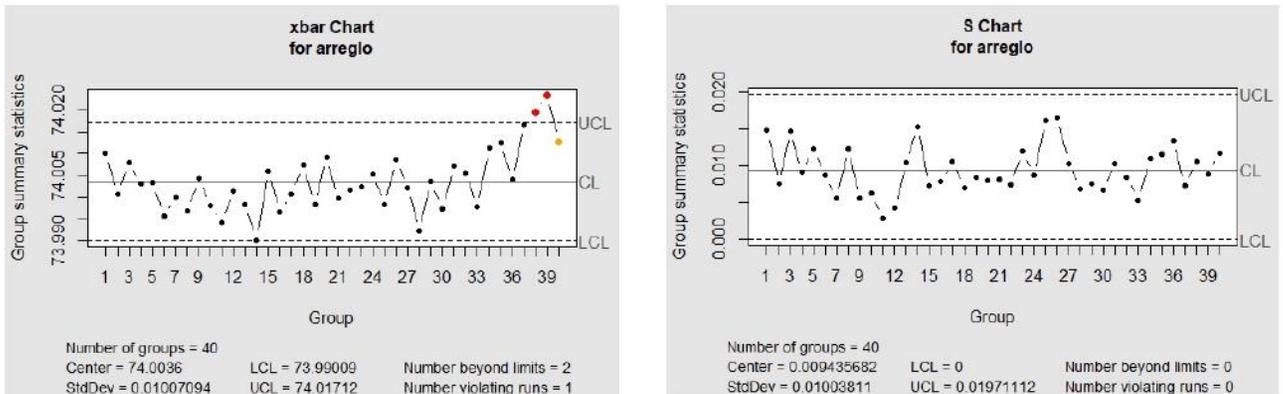


Figura 3: Cartas de control iniciales.

Se pueden apreciar en la carta  $\bar{X}$  que existen dos puntos fuera de control y uno como corrida. Si bien es un inconveniente importante, son requeridos estos puntos para evidenciar el funcionamiento de la función `qcc.options` que será explicada a continuación.

### Función `qcc.options`

La función `qcc.options` permite establecer parámetros iniciales en el comportamiento de la librería `qcc`. Si se llama a la función de la forma `qcc.options()` se desplegarán en la consola los valores de todos los parámetros con sus valores por defecto como se puede evidenciar en la *Figura 4*.

El nuevo código a ejecutar se muestra a continuación:

```
qcc.options(run.length=7)
qcc.options('violating.runs' = list(pch = 17, col = 'blue'))
qcc.options('beyond.limits' = list(pch = 17, col = 'red'))
qcc.options(bg.margin='white',bg.figure='grey')
qcc.options(cex=1,font.stats=4,cex.stats=0.8)
```

Los parámetros que corresponden a esta función se clasifican en:

```

> qcc.options()
$exp.R.unscaled
 [1] NA 1.128 1.693 2.059 2.326 2.534 2.704 2.847 2.970 3.078 3.173 3.258
[13] 3.336 3.407 3.472 3.532 3.588 3.640 3.689 3.735 3.778 3.819 3.858 3.895
[25] 3.931

$se.R.unscaled
 [1] NA 0.8525033 0.8883697 0.8798108 0.8640855 0.8480442 0.8332108
 [8] 0.8198378 0.8078413 0.7970584 0.7873230 0.7784873 0.7704257 0.7630330
[15] 0.7562217 0.7499188 0.7440627 0.7386021 0.7334929 0.7286980 0.7241851
[22] 0.7199267 0.7158987 0.7120802 0.7084528 0.7050004 0.7017086 0.6985648
[29] 0.6955576 0.6926770 0.6899137 0.6872596 0.6847074 0.6822502 0.6798821
[36] 0.6775973 0.6753910 0.6732584 0.6711952 0.6691976 0.6672619 0.6653848
[43] 0.6635632 0.6617943 0.6600754 0.6584041 0.6567780 0.6551950 0.6536532
[50] 0.6521506

$beyond.limits
$beyond.limits$pch
 [1] 19

$beyond.limits$col
 [1] "red"

$violating.runs
$violating.runs$pch
 [1] 19

$violating.runs$col
 [1] "orange"

```

Figura 4: Resultado del comando `qcc.options()` en la consola.

## Argumentos estadísticos

- ✓ **exp.R.unscaled:** Un vector el cual especifica para cada tamaño de muestra, el valor estimado del rango relativo  $R/\sigma$  para una distribución normal. Estos valores aparecen en las tablas de factores para construir cartas de control como  $d2$ . Si se utiliza `qcc.options()` aparecerán en la consola en el mismo orden que lo evidencia en la tabla de consulta. Este argumento se recomienda no modificar debido a que ya son datos establecidos para poder calcular los límites de control.
- ✓ **se.R.unscaled:** Funciona de manera similar al argumento anterior, la diferencia radica en que este especifica para cada tamaño de muestra el error estándar del rango relativo  $R/\sigma$  para una distribución normal. Este se encuentra en las tablas como  $d3$  y de la misma manera, si se usa el comando `qcc.options()` aparecerán los valores aproximados de las tablas de factores en la consola. Para este se hace la misma recomendación de no modificar.
- ✓ **run.length:** Especifica un número máximo de puntos que estén por debajo o encima de la línea central consecutivos y una vez alcance este valor, a partir del valor indicado señalará estos puntos como corridas. El número establecido es inversamente proporcional al número de corridas que detectará *Rstudio* en el gráfico de control, es decir, entre más elevado sea el número, menores serán los puntos de corrida que el software grafique. El valor por defecto es siete, pero bien puede declarar de la forma `qcc.options=7`

## Figuras, colores y tipografía

- ✓ **pch:** Hace parte de la función `pars`, y este a su vez se relaciona con la función `points`. La idea de este argumento es indicar la forma o el símbolo que deben tener los puntos en un gráfico. Existe una gran variedad de símbolos sin embargo los más utilizados se muestran en la *Figura 5* la cual especifica todos los valores correspondientes del cero hasta 25 y su respectivo símbolo.



Figura 5: Valores y su correspondiente símbolo para el argumento *pch*.

- ✓ **beyond.limits\$pch:** Este argumento cambia el simbolo para los puntos que están fuera de los límites de control.
- ✓ **violating.runs\$pch:** Por otro lado, este argumento permite modificar el simbolo de los puntos en una corrida.

En cuanto se trata de colores, la función **qcc.options** permite realizar una serie de cambios concisos a ciertos aspectos. Estos son:

- ✓ **beyond.limits\$col:** Cambia el color a los puntos que están fuera de los límites de control.
- ✓ **violating.runs\$col:** Modifica el color de los puntos que se encuentra en una sucesión o corrida.

De acuerdo a la información dada, ejecute las líneas:

```
qcc.options('violating.runs' = list(pch = 17, col = 'blue'))
qcc.options('beyond.limits' = list(pch = 17, col = 'red'))
```

- Los nombres “violating.runs” y “beyond.limits” indican respectivamente si la modificación se realizará a los puntos que tengan corridas y a los que se encuentre fuera de control.
- La función *list* permite agrupar los argumentos **pch** y **col** en una sólo línea.

De esta manera, se pretende que los puntos corridos y fuera de control tengan una figura triangular, y respectivamente sean coloreados con azul y rojo.

- ✓ **bg.margin:** Este argumento cambia el color de la parte externa del gráfico. Declare:

```
qcc.options(bg.margin='white')
```

Con esto la parte externa tendrá un tono blanco.

- ✓ **bg.figure:** Permite modificar el tono de la parte interna de la figura. En el caso de las cartas de control, es el área donde se encuentran todos los puntos. Especifique `qcc.options(bg.margin='grey')` para que la cartas de control en su interior tengan un tono gris.
- ✓ **cex:** Un valor numérico que permite cambiar el tamaño de las proporciones del gráfico de control. Este valor indica un aumento de forma porcentual en base al tamaño por defecto, es decir, si el valor por defecto es **1**, indica que se mantendrá el tamaño base, por otra parte, la cifra **0.7** indica que las proporciones en el gráfico tendrán una reducción del 30%. Para este caso dictamine el valor por defecto.

- ✓ **font.stats:** Permite cambiar la tipografía a los datos estadísticos que se encuentran en la parte inferior del gráfico. Este argumento funciona de la misma manera y con los mismos parámetros que el argumento **font** el cual se utilizó en la práctica 5 correspondiente al *Diagrama de Ishikawa*. Se muestran nuevamente la tabla con los parámetros establecidos:

Tabla 1: Parámetros de tipografía para el argumento *font.stats*

Dígito	Tipografía	Característica
1, 18 y 20 en adelante	Arial MT	Sin negrita ni cursiva
2		Negrita
3		Cursiva
4		Negrita y cursiva
5	Alfabeto griego	Sin negrita ni cursiva
6	Perpetua	Sin negrita ni cursiva
7		Negrita
8		Cursiva
9		Negrita y cursiva
10	Typewriter	Sin negrita ni cursiva
11		Negrita
12		Cursiva
13		Negrita y cursiva
14	Century Gothic	Sin negrita ni cursiva
15		Negrita
16		Cursiva
17		Negrita y cursiva
19	Symbol	Sin negrita ni cursiva

En este caso se elegirá el valor 4 que corresponde la fuente *Arial MT* en negrita y cursiva.

- ✓ **cex.stats:** Realiza la misma función que el argumento **cex**, pero solamente para los datos estadísticos de la parte inferior de los gráficos de control. Especifique para este caso un valor de 0.8 de la forma `qcc.options(cex.stats=0.8)`.

! Para regresar estos parámetros a sus estados por defecto, tendrá que sacar de escena el paquete *qcc*. Esto se logra yendo al entorno de utilidades, luego haciendo click en el botón “packages” y desactivando la casilla marcada con el nombre del paquete y volviendolo a activar.

De esta manera, si vuelve a ejecutar los comandos para generar las cartas de control `qcc(arreglo, type='xbar')` y `qcc(arreglo, type='S')` respectivamente, los gráficos resultantes haciendo uso de la función *qcc.options* tendrán la siguiente forma:

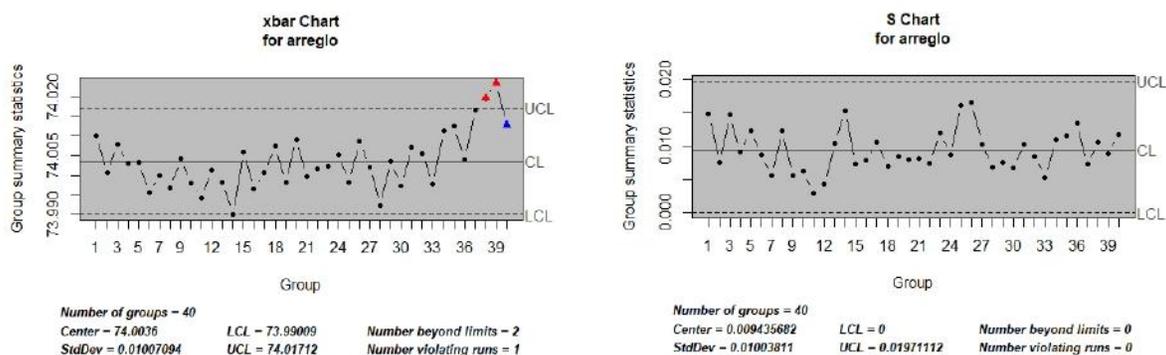


Figura 6: Cartas de control  $\bar{X}$  y  $S$  de acuerdo a las modificaciones realizadas en la función `qcc.options`.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuesta deben estar relacionadas con lo observado durante la práctica.

A partir del siguiente problema resuelva los incisos planteados:

Tabla 2: Datos de altura de llenado.

Número de muestra	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
1	2.5	0.5	2.0	-1.0	1.0	-1.0	0.5	1.5	0.5	-1.5
2	0.0	0.0	0.5	1.0	1.5	1.0	-1.0	1.0	1.5	-1.0
3	1.5	1.0	1.0	-1.0	0.0	-1.5	-1.0	-1.0	1.0	-1.0
4	0.0	0.5	-2.0	0.0	-1.0	1.5	-1.5	0.0	-2.0	-1.5
5	0.0	0.0	0.0	-0.5	0.5	1.0	-0.5	-0.5	0.0	0.0
6	1.0	-0.5	0.0	0.0	0.0	0.5	-1.0	1.0	-2.0	1.0
7	1.0	-1.0	-1.0	-1.0	0.0	1.5	0.0	1.0	0.0	0.0
8	0.0	-1.5	-0.5	1.5	0.0	0.0	0.0	-1.0	0.5	-0.5
9	-2.0	-1.5	1.5	1.5	0.0	0.0	0.5	1.0	0.0	1.0
10	-0.5	3.5	0.0	-1.0	-1.5	-1.5	-1.0	-1.0	1.0	0.5
11	0.0	1.5	0.0	0.0	2.0	-1.5	0.5	-0.5	2.0	-1.0
12	0.0	-2.0	-0.5	0.0	-0.5	2.0	1.5	0.0	0.5	-1.0
13	-1.0	-0.5	-0.5	-1.0	0.0	0.5	0.5	-1.5	-1.0	-1.0
14	0.5	1.0	-1.0	-0.5	-2.0	-1.0	-1.5	0.0	1.5	1.5
15	1.0	0.0	1.5	1.5	1.0	-1.0	0.0	1.0	-2.0	-1.5

El volumen de llenado de las botellas de refresco es una característica de la calidad importante. El volumen se mide (aproximadamente) colocando un medidor sobre la boca de la botella y comparando la altura del líquido en el cuello de la botella con una escala codificada. En esta escala, una lectura cero corresponde a la altura de llenado correcta [1]. Se analizan 15 muestra de tamaño  $n = 10$  y las altura de llenado se muestran en la Tabla 2:

1. Establecer las cartas de control  $\bar{X}$  y  $S$  para este proceso. ¿El proceso muestra control estadístico? De ser necesario, construir los límites de control revisados.

2. Establecer una carta  $R$  y compararla con la carta  $S$  del inciso 1.
3. Establecer una carta  $S^2$  y compararla con la carta  $S$  del inciso 1.

## 6. Referencias

- [1] D. V. Montgomery, Douglas C y Verbeeck. *Control estadístico de la calidad*. Limusa Wiley, 2004.

# PRACTICA 8: CARTA DE CONTROL X-ONE PARA MEDICIONES INDIVIDUALES.

## 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control de mediciones individuales *X one* mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente gráficos de control de rango móvil a partir de una carta *R* mediante el uso del paquete *Quality Control Charts (qcc)*.
- ✓ Interpretación y análisis de la información a partir de los gráficos resultantes.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

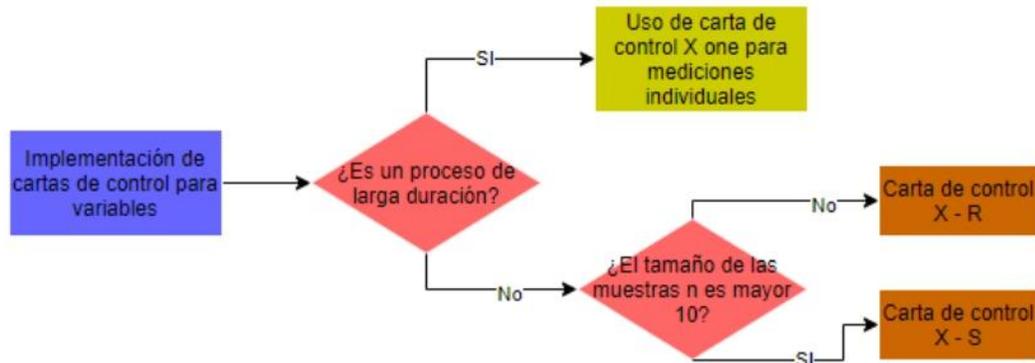


Figura 1: Diagrama de flujo correspondiente a la hora de elegir la carta de control para mediciones individuales:

### 3.1 Definición.

La carta de control *X one* para mediciones individuales hace parte parte del grupo de cartas de control para variables. Se basa en como su nombre lo indica, hacer el respectivo análisis a cada uno de los productos resultantes. Esto se toma en cuenta cuando los procesos son demasiado lentos, en los cuales se puede tardar un día o más para tener una muestra [1]. Algunos casos pueden ser:

- ✓ Productos químicos.

- ✓ Industria de licores.
- ✓ Variables administrativas como índices de productividad, consumo de servicios públicos.

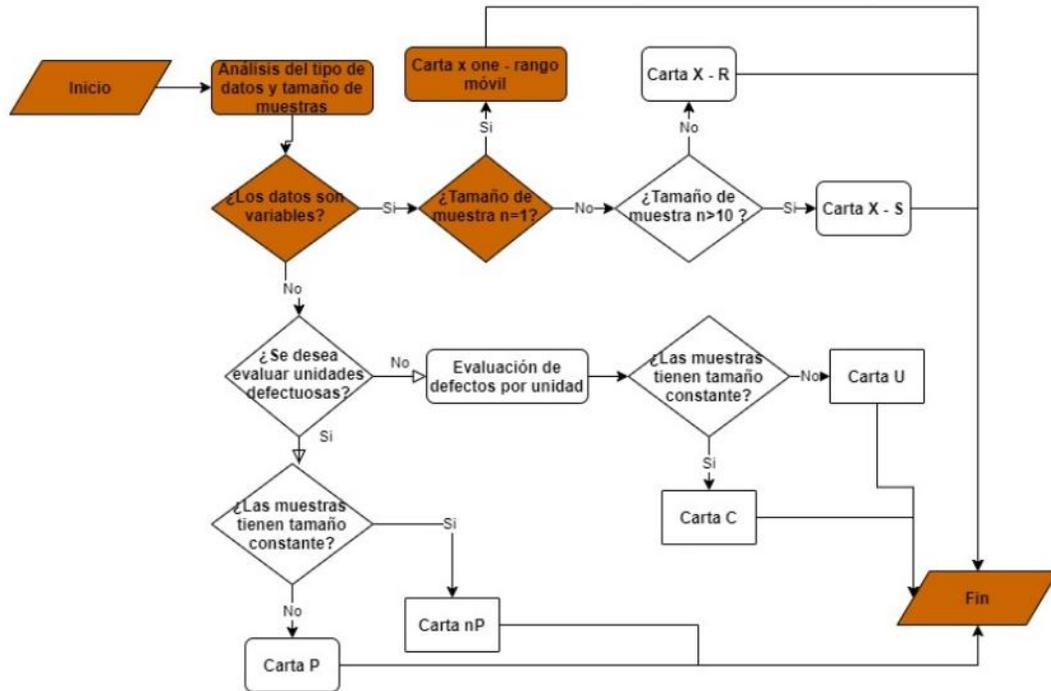


Figura 2: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta  $X - One$  para mediciones individuales.

Para poder elaborar este gráfico, los límites de control respectivamente son:

$$UCL = \bar{X} + 3\left(\frac{\bar{R}}{d_2}\right)$$

$$CL = \bar{X}$$

$$LCL = \bar{X} - 3\left(\frac{\bar{R}}{d_2}\right)$$

! El valor de la constante  $d_2$  lo puede encontrar en la *Tabla de valores para la construcción de cartas de control* disponible en libros de control estadístico de calidad o bien en internet. Tenga presente el correspondiente valor para el tamaño de la muestra especificado.

#### 4. Montaje y procedimiento

**Docente:** Indique al estudiante que abra el software *R studio* y que elabore una carta de control de mediciones individuales  $X one$  y una de rango móvil a partir de los datos del ejercicio planteado. A continuación,

realice la explicación paso a paso del código correspondiente a la elaboración de los gráficos.

**Estudiante:** Siga las instrucciones del docente detalladamente en pro de la comprensión a la hora de elaborar la carta de mediciones individuales *X one*. Para lograr este objetivo, se partirá del siguiente caso:

### Planteamiento del problema.

Tabla 1: Temperatura y rango móvil para horneado de impresiones en láminas de acero.[1]

Muestra	Temperatura en el horno	Rango móvil	Muestra	Temperatura en el horno	Rango móvil
1	125.1	0	13	126.4	1.6
2	127.5	2.4	14	128.3	1.9
3	122.7	4.8	15	129.5	1.2
4	126.4	3.7	16	128.1	1.4
5	125.5	0.9	17	125.1	3.0
6	130.5	5.0	18	128.5	3.4
7	127.3	3.2	19	125.0	3.5
8	127.5	0.2	20	126.3	1.3
9	127.3	0.2	21	126.5	0.2
10	123.0	4.3	22	127.9	1.4
11	123.5	0.5	23	129.5	1.6
12	128.0	4.5	24	131.9	2.4

En una empresa se hacen impresiones en láminas de acero, que posteriormente se convierten en recipientes de productos de otras empresas. Un aspecto importante a vigilar en dicha impresión es la temperatura de “horneada” que, entre otras cosas, adhiere y seca la pintura una vez que ésta se imprimió. En una fase particular de la horneada se tiene que la temperatura de cierto horno debe ser  $125^{\circ}C$ , con una tolerancia de  $\pm 5^{\circ}C$  si no se cumple con tal rango de temperatura, entonces se presentan problemas en la calidad final de la impresión.

Para controlar de manera adecuada tal característica de proceso se usa una carta de control, y evidentemente aquí no tiene sentido tomar una muestra de tamaño cuatro, ya que si se miden las temperaturas del horno de manera consecutiva, entonces éstas serán prácticamente las mismas; y si las mediciones no se hacen de manera consecutiva y se hacen, por ejemplo, cada hora, entonces habrá que esperar cuatro horas para graficar un punto en la carta de control y así analizar el proceso. De este modo resulta más práctico hacer, periódicamente, una medición de la temperatura en el horno, y en cuanto se obtiene el dato graficarlo en la carta. Es decir, **resulta más práctico llevar una carta de control para mediciones individuales**[1].

### Elaboración del código.

Se presenta en la *Tabla 1* los datos correspondientes a las 24 muestras tomadas para cada temperatura y su respectivo rango móvil. A continuación se evidencia la estructura del código a ejecutar y se explica debajo de este el paso a paso:

! los puntos suspensivos que evidenciará en los vectores “temperatura” y “rango\_movil” son sólo una forma de acotar el código. Para que este funcione deberá colocar todos los valores de la columna correspondiente que se encuentran en la *Tabla 1*.

```
temperatura<-c(125.1, 127.5, 122.7, 126.4,...,131.9)
rango.movil<-c(0, 2.4, 4.8, 3.7, 0.9,...,2,4)
datos<-data.frame(temperatura,rango.movil)
attach(datos)
datos
library(qcc)
```



Recuerde que si no se encuentra instalado el paquete *qcc* en el terminal, podrá hacerlo ejecutando la línea `install.packages("qcc")` en el área de edición de scripts, o bien, buscándolo en el entorno de utilidades haciendo click en el botón “Packages” y enseguida en “Install”.

1. Elabore en *Rstudio* dos vectores que contengan en orden los datos de las columnas *Temperatura en el horno* y *Rango móvil* y nómbrelas como “temperatura” y “rango\_movil” respectivamente.
2. Elabore un *dataframe* con los vectores generados en el inciso anterior con el nombre de “datos” haciendo uso de la función `data.frame` en la cual colocará dentro de paréntesis los nombres de los vectores. De tal forma el código a elaborar tiene la forma:

```
datos<-data.frame(temperatura,rango_movil)
```

3. Al llamar el dataframe elaborado por su nombre podrá apreciar el arreglo correspondiente en la consola.
4. Haga uso de la función `attach` para trabajar con las columnas generadas del dataframe *datos*.
5. Haga la puesta en escena de la librería *qcc* mediante función `library`.

Para la elaboración de la carta *X-one* deberá hacer uso de la función `qcc`. Recuerde que la forma general de la función tiene la siguiente estructura:



Cuando se habla la forma general de una función, esto sólo indica la forma correcta de su uso y el tipo de argumentos que esta utiliza (lógico, numérico, vector, entre otros). **No se debe ejecutar esta línea.**

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

A continuación se muestra la línea de código **la cual si debe ser ejecutada**:

```
qcc(temperatura, type = "xbar.one",title = "Carta X - one de mediciones individuales
para temperatura del horno",std.dev="MR")
```



No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

### Otros argumentos de la función `qcc`.

Deberá realizar los siguientes ajustes a los argumentos descritos a continuación:

- ✓ **data:** Se tomarán los valores del vector “temperatura”. Al haber hecho uso de la función `attach` solamente tendrá que colocar el nombre de esta.
- ✓ **type:** Para este argumento deberá especificar “`xbar.one`” para dar la indicación a *Rstudio* de que debe elaborar la carta para mediciones individuales.
- ✓ **title:** La gráfica generada tendrá como título “Carta X - one de mediciones individuales para temperatura del horno.”

Los siguientes argumentos se pueden modificar, sin embargo **se recomienda dejar sus valores por defecto**. A continuación se pone de manifiesto su correcto uso:

- ✓ **std.dev:** Un valor o un método el cual dictamina la desviación estándar de la carta de control. Los métodos que dispone *Rstudio* para calcular la desviación estándar son:
  - **“UWARE-R”:** Promedio no ponderado de subgrupos estimados basado en rangos.
  - **“UWARE-SD”:** Promedio no ponderado de subgrupos estimados basados en desviaciones estándar.
  - **“MVLUE-R”:** Estimador insesgado de varianza mínima calculado como un promedio ponderado de subgrupos aproximados basado en rangos.
  - **“MVLUE-S”:** Estimador insesgado de varianza mínima calculado como un promedio ponderado de subgrupos aproximados basado en desviaciones estándar.
  - **“RMSDF”:** Estimador de la raíz de la media calculado como un promedio ponderado de subgrupos aproximados basado en subgrupos de desviaciones estándar.

En la siguiente tabla se muestra de manera compacta los métodos disponibles para los tipos de gráficos “`xbar`”, “`R`” y “`S`”:

*Tabla 2: Métodos disponibles de desviaciones estandar.*

Método	“ <code>xbar</code> ”	“ <code>R</code> ”	“ <code>S</code> ”
“UWARE-R”	Por defecto	Por defecto	No disponible
“UWARE-SD”	Disponible	No disponible	Por defecto
“MVLUE-R”			No disponible
“MVLUE-S”		No disponible	Disponible
“RMSDF”		No disponible	

La carta de control para mediciones individuales *X one* cuenta con sus propios métodos para el cálculo de la desviación estándar. Estos son:

- **“MR”:** Rango móvil. Esta desviación es la que se encuentra estipulada por defecto.
- **“SD”:** Desviación estándar de la muestra.

En ese orden de ideas, establezca para la carta de mediciones individuales la desviación “**MR**” de la forma `std.dev= ‘MR’`

! Si no establece un tipo de desviación estándar *Rstudio* dictaminará un método de desviación por defecto expuesto en este inciso para la carta con la que esté trabajando.

- ✓ **confidence.level:** Es un valor que oscila entre 0 y 1 el cual indica el nivel de confianza del cálculo de los límites de control. En e
- ✓ **nsigmas:** Número de sigmas especificado para los límites de control. Por defecto será un valor de 3. Este argumento acepta cualquier valor mayor a cero, sin embargo, lo recomendable es utilizar un valor de sigma de tres.

De esta manera el gráfico resultante se presenta en la *Figura 3*:

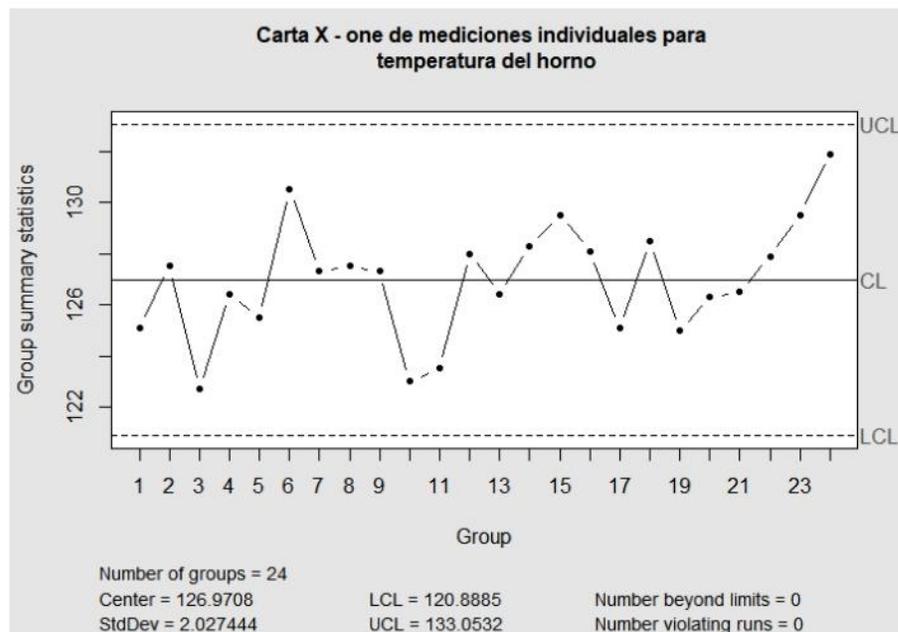


Figura 3: Carta de mediciones individuales.

Si bien la carta no presenta puntos fuera de control, hay que tomarla con pinzas ya que, se puede evidenciar un comportamiento al alza desde el punto 19 hasta el 24, lo cual indica la existencia de una causa asignable que esté haciendo estos aumentos de temperatura.

Por otra parte, a pesar de que la carta de mediciones individuales muestra un panorama reelevantante del proceso, existe una carta que complementa su función: **la carta de rango móvil**. Este gráfico funciona de una manera similar a la carta de rangos, la diferencia radica en los estadísticos  $D_3$  y  $D_4$  que tendrán tamaño 2 (Esto a que cada dato del rango móvil se obtiene de su consecutivo[1]). De tal manera, para este caso los límites de control se calcularán de la forma:

$$\begin{aligned}
 UCL &= D_3 \bar{R} = 0(2.287) = 0 \\
 CL &= \bar{R} = 2.2287 \\
 LCL &= D_4 \bar{R} = 3.2686(2.287) = 7.475
 \end{aligned}$$

Debido a que se planea realiza una carta de *rango móvil* a partir del type “R” se tendrán que colocar los límites de control de forma manual. Para lograr esto, haga uso de la función `qcc`. Recuerde nuevamente la forma general de esta función:

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

De esta manera la nueva línea de código que **debe ejecutar** corresponde a la siguiente estructura y a la postre se explica el respectivo concepto de cada argumento:

```
qcc(rango_movil,type = ‘R’,sizes = 2, limits = c(0,7.475),center = 2.287, title =
‘Carta de rango móvil para datos de temperatura del horno’)
```

Como puede apreciar en el código, esta vez en la nueva línea escoga en el argumento **data** el vector “rango\_movil” y seleccione en **type** el tipo “R”. Luego de esto realice los ajustes pertinentes a los siguientes argumentos:

- ✓ **sizes**: Valor que indica el tamaño de las muestras. Para el caso del rango móvil será de dos, por lo que dictamine de la forma `sizes = 2`.
- ✓ **limits**: Vector que permite establecer los valores para los límites de control superior e inferior respectivamente. Para esto declare los datos calculados de los límites de la forma `limits = c(0,7.475)`.
- ✓ **center**: Establece el valor de la línea central (2.2287 para este caso). Realice `center = 2.287` para tal fin.
- ✓ **title**: Establezca “Carta de rango móvil para datos de temperatura del horno” como título para esta carta de control.

De esta manera, la carta de control para rango móvil se puede apreciar en la *Figura 4*:

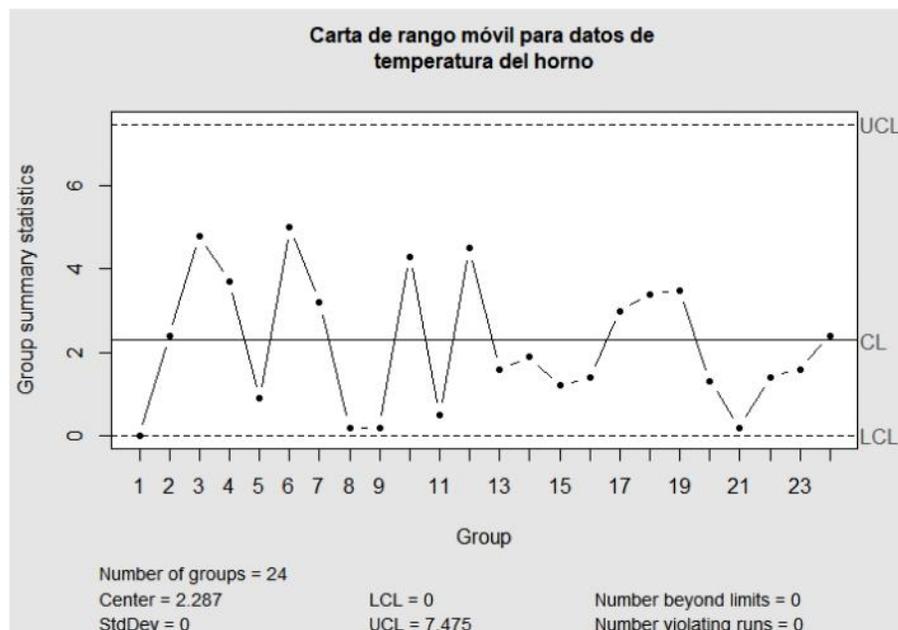


Figura 4: Carta de rango móvil correspondiente.

! Recuerde que puede hacer un espacio de línea horizontal para los títulos haciendo uso de una barra invertida junto la letra n.

Este gráfico de control permite medir la variabilidad del proceso, sin embargo hay muchos estudios que muestran que la carta de mediciones individuales es herramienta suficiente para poder detectar cambios en la media y en la variabilidad del proceso.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

A partir de la información suministrada resuelva los siguientes puntos:

*En una empresa, los gastos por consumo de agua son considerablemente altos, por lo que se decide establecer una carta de mediciones individuales con el propósito de detectar anomalías y buscar reducir el consumo. Se tomará la lectura al medidor de agua todos los lunes para cuantificar el consumo de la semana anterior. Los datos (en m<sup>3</sup>) obtenidos en 20 semanas se muestran en la Tabla 2:*

Tabla 3: Consumo de agua para 20 semanas.

Semana	Consumo
1	562
2	577
3	536
4	536
5	650
6	525
7	533
8	569
9	563
10	553
11	549
12	480
13	557
14	555
15	518
16	557
17	553
18	540
19	517
20	571

1. ¿Por qué no es apropiado analizar estos datos mediante una carta  $\bar{X} - R$  ?
2. Mediante una carta para mediciones individuales y una carta de rangos móviles, investigue si el consumo de agua estuvo en control estadístico.

3. Explique de manera clara los límites de control que obtenga.
4. Obtenga los límites naturales del proceso.

## 6. Referencias

- [1] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010

## PRACTICA 9: CARTA PARA PROPORCIÓN DE ARTÍCULOS DEFECTUOSOS P

### 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control para proporción de artículos defectuosos  $p$  mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Desarrollar los conocimientos necesarios para elaborar cartas  $p$  con límites constantes (límite promedio) y variables.
- ✓ Establecer los conocimientos necesarios que permitan añadir nuevos puntos a un gráfico de control ya establecido.

### 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <b>R</b> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <b>R Studio</b> .....	1

### 3. Teoría relacionada

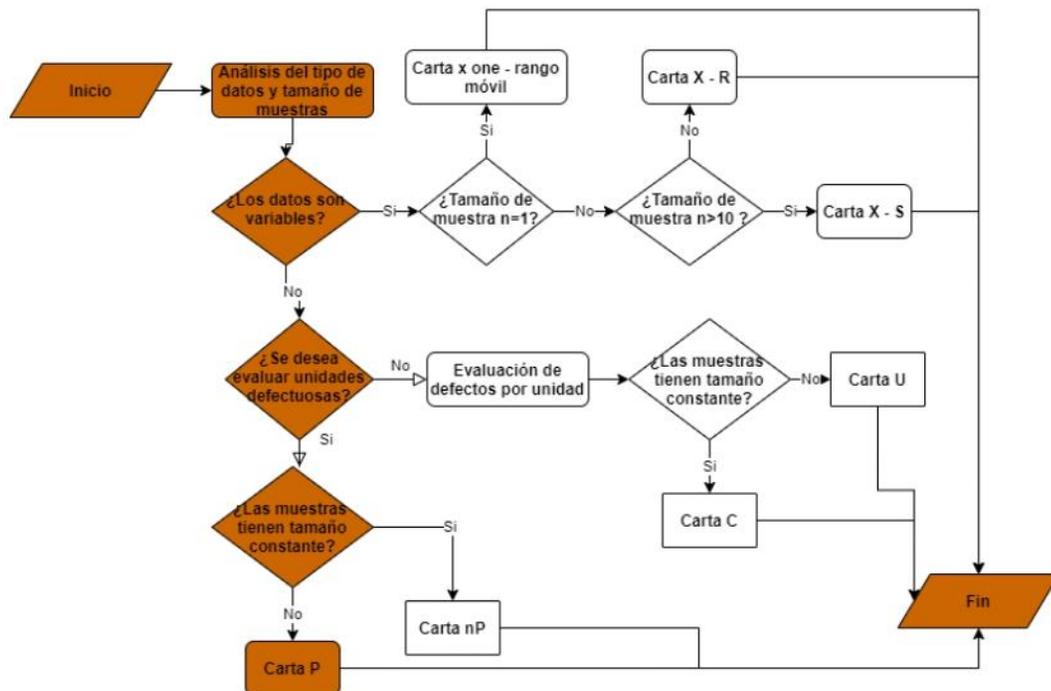


Figura 1: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta  $p$ .

### 3.1 Definición.

Las cartas de control por atributos son la otra clasificación para el análisis estadístico por medio de cartas de control. Estas se desarrollan con el fin evaluar procesos en los que no es posible cuantificar mediante una magnitud si no mediante parámetros tales como *conforme* o *no conforme* (por ejemplo, al cuantificar la magnitud de la rayadura de una pieza, o el número de defectos que esta cuenta).

La carta para proporción de artículos defectuosos o más conocida como *carta p* permite establecer una proporción  $p_i$  de artículos defectuosos  $d_i$  en un tamaño determinado de muestra o lote  $n_i$  de tal manera que la proporción se obtiene a partir de:

$$p_i = \frac{d_i}{n_i} \quad (1)$$

#### Cálculo de los límites de control.

A diferencia de las cartas de control para variables, los límites de control en estos gráficos no dependen de una constante, si no más bien en el tamaño de la muestra  $n_i$  para cada proporción. Siguiendo una distribución binomial [1], se tiene que los límites de control son

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (2)$$

$$CL = \bar{p} \quad (3)$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (4)$$

#### Variaciones de la carta P.

Existen casos en los que el tamaño de la muestra  $n_i$  es variable (es decir, no todas las muestras tendrán el mismo tamaño). Con lo cual se plantean dos alternativas:

- ✓ Utilizar el tamaño promedio de las muestras  $\bar{n}$ . Esto es recomendable cuando el número de muestras no sobrepase el 20% del total del tamaño de estas.
- ✓ Utilizar límites de control variables, es decir, calcular para cada tamaño de muestra  $n_i$  los límites de control respectivos sin embargo hay algunos casos en los que esto puede resultar inconveniente a la hora de evaluar la carta de control. Una alternativa extra para este caso es utilizar la **carta p estandarizada**[1], la cual utiliza como UCL y LCL los valores de 3 y -3 respectivamente. La proporción promedio  $\bar{p}$  se corresponde con los anteriores y cada proporción  $p_i$  se somete a la ecuación:

$$Z_i = \frac{p_i - \bar{p}}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \quad (5)$$

## 4. Montaje y procedimiento

### Carta $p$ con límites constantes (tamaño promedio de las muestras).

**Docente:** Indique al estudiante que abra el software *Rstudio* y que elabore una carta para proporción de artículos defectuosos o *Carta  $p$*  con límites constantes a partir de los datos del ejercicio planteado. A continuación, realice la explicación paso a paso de las funciones y del código correspondiente a la elaboración de los gráficos.

**Estudiante:** Siga las instrucciones del docente detalladamente en pro de la comprensión a la hora de elaborar esta carta  $p$ . Para lograr este objetivo, se partirá del siguiente caso:

### Planteamiento del problema.

Tabla 1: datos obtenidos durante una semana para cierto tipo de válvulas

Muestra o lote	Tamaño de lote	Artículos defectuosos	Proporción
1	300	15	0.05
2	300	12	0.04
3	300	15	0.05
4	300	7	0.02
5	330	16	0.05
6	300	6	0.02
7	300	18	0.06
8	280	10	0.04
9	290	9	0.03
10	300	25	0.08
11	300	9	0.03
12	300	4	0.01
13	300	7	0.02
14	300	9	0.03
15	305	5	0.02
16	295	15	0.05
17	300	19	0.06
18	300	7	0.02
19	300	12	0.04
20	300	10	0.03
21	300	4	0.01

En una empresa del ramo metal-mecánico se fabrican válvulas. Después del proceso de fundición se hace una inspección y las piezas que no cumplen con ciertas características se rechazan. Las razones por las que pueden rechazarse son diversas: piezas incompletas, porosas, mal formadas, etc. Para evaluar la variabilidad y la magnitud de la proporción de piezas defectuosas en el proceso de fundición se decide implementar una carta  $p$ .

Aunque regularmente el tamaño de lote es fijo,  $n = 300$ , en ocasiones, por diferentes motivos, en algunos lotes se hacen unas cuantas piezas de más o de menos, como se aprecia en la Tabla 1. Para calcular los límites se usará el tamaño de muestra (lote) promedio.

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

## Elaboración del código.

**Docente:** Realice la explicación paso a paso de las funciones y del código correspondiente a la elaboración del gráfico de control.

**Estudiante:** Importe los datos de la *tabla 1* a *Rstudio* como **un archivo de Excel** con el nombre de *Datos\_carta\_p*.

! Recuerde que puede hacer la importación en la ventana superior derecha del software, luego dando click en el botón “Import Dataset” y luego dando click en “From Excel...”

Enseguida, ejecute paso a paso el código que se se explica de forma detallada a continuación para la elaboración de la carta de control  $p$  cuyo resultado se evidencia en la *Figura 2* y su procedimiento debajo de la imagen:

! Se recomienda al momento de elaborar los datos de la *Tabla 1* separar los títulos con un guión al piso en vez de un espacio.

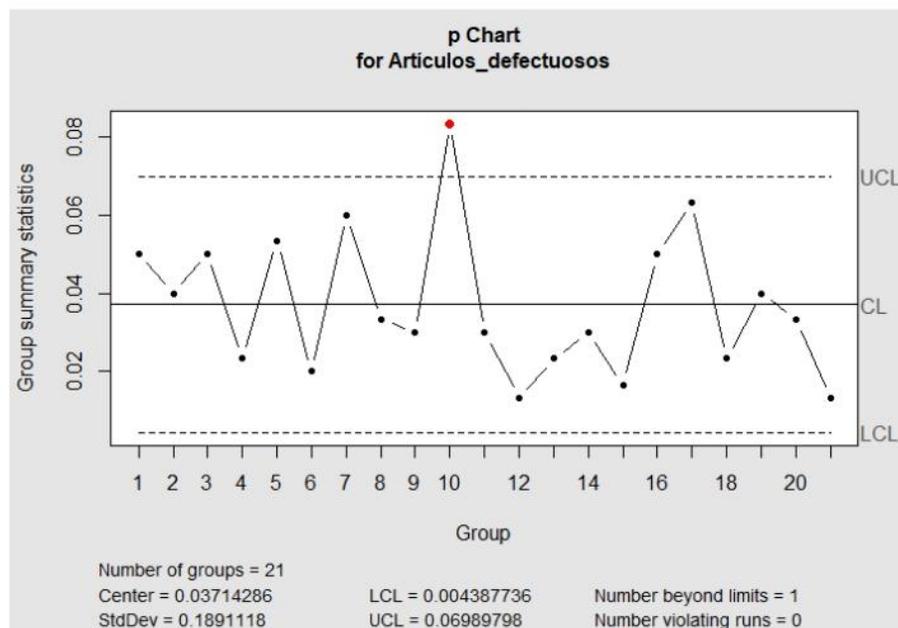


Figura 2: Carta  $p$  para proporción de artículos defectuosos en válvulas.

! Recuerde que si no se encuentra instalado el paquete *qcc* en el terminal, podrá hacerlo ejecutando la línea `install.packages('qcc')` en el área de edición de scripts, o bien, buscándolo en el entorno de utilidades haciendo click en el botón “Packages” y enseguida en “Install”.

El código a elaborar presenta la siguiente estructura:

```
library('qcc')  
attach(Datos_carta_p)  
Datos_carta_p
```

```
qcc(Artículos_defectuosos,type = 'p',sizes = 300)
```

El paso a paso del código de evidencia a continuación:

1. Realice la puesta en escena de la librería *qcc*.
2. Haga uso de la función `attach` para los datos que ha importado al software anteriormente.
3. Realice un chequeo de los datos en la consola simplemente ejecutando el nombre de esta en el área de edición de scripts.
4. Haga uso de la función `qcc`. Recuerde que la forma general de esta función corresponde a la siguiente estructura:

! Cuando se habla la forma general de una función, esto sólo indica la forma correcta de su uso y el tipo de argumentos que esta utiliza (lógico, numérico, vector, entre otros). **No se debe ejecutar esta línea.**

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

Deberá dictaminar en el argumento `data` el nombre de la tercera columna de los datos que se encuentran en la *Tabla 1* (“Artículos Defectuosos”); en `type` escriba entre comillas “p” para obtener esta carta de control. Finalmente haga uso del argumento `sizes` dando 300 como dimensión para todas las muestras (esto debido a que este es el valor promedio del tamaño de las muestras).

### Añadir puntos a la carta P.

Una vez elaborado el gráfico de control inicial, se dispone del siguiente escenario en la que **se deberán añadir nuevos puntos a la carta p**:

*Dado que el nivel promedio de defectos del proceso se consideró demasiado alto, se decidió realizar un proyecto de mejora siguiendo el ciclo PHVA. Después de aplicar las soluciones propuestas para reducir la proporción de piezas defectuosas, se tiene la siguiente proporción de defectuosos en 19 lotes de tamaño 300 cada uno:*

0.013, 0.010, 0.030, 0.020, 0.007, 0.017, 0.010, 0.027, 0.000, 0.017, 0.040, 0.013, 0.017, 0.013, 0.017, 0.013, 0.020, 0.030, 0.007.

**Docente:** Realice la explicación respectiva para agregar nuevos puntos a la carta de control partiendo de la base de los 19 datos mencionados en el enunciado anterior. A la postre, realice la didáctica respectiva de los argumentos requeridos.

**Estudiante:** Elabore un vector con los 19 nuevos datos y nexos, multiplique este por 300 (tenga presente que aún se pretende elaborar una carta con límites constantes) para obtener el número de defectos para cada muestra. De acuerdo a la ecuación (1) para obtener el número de defectos  $d_i$  se debe multiplicar la proporción  $p_i$  que es este caso son los datos de los vectores, por el tamaño de cada lote  $n_i$  en donde 300 es el valor para cada uno. En ese orden de ideas, deberá multiplicar el vector generado (en este caso nombrado como “nuevo”) al final usando un asterisco (\*) por el valor 300 tal como se evidencia en la *Figura 3*. Para poder apreciar el resultado puede ejecutar el nombre de este vector en la consola:

```

> nuevo<-c(0.013, 0.010, 0.030, 0.020, 0.007, 0.017, 0.010,
+         0.027, 0.000, 0.017, 0.040, 0.013, 0.017, 0.013,
+         0.017, 0.013, 0.020, 0.030, 0.007)*300
> nuevo

```

Figura 3: Creación del vector correspondiente a los 19 nuevos datos

Haga uso del mismo comando que utilizó para elaborar la carta de la *Figura 1*. Luego deberá realizar algunos ajustes a los siguientes argumentos que se explican a continuación teniendo presente este nuevo código:

```

qcc(Artículos_defectuosos,type = 'p', sizes = 300,data.name='antiguos datos',
newdata=nuevo, newsizes = 300,newdata.name='datos ciclo PHVA')

```

- ✓ **data.name:** Cadena de caracteres que dictamina el nombre del primer conjunto de datos. Si no se asigna nada este será nombrado por defecto como se especificó el conjunto de datos en el argumento **data**. Escriba el argumento de la forma **data.name='antiguos datos'** para la nueva carta de control.
- ✓ **newdata:** Permite agregar un conjunto de datos a las cartas de control que no se había establecido desde un inicio. Este puede tener la forma de matriz, vector o dataframe. Para ello, realizará el código de la forma **newdata=nuevo**. Con esto serán añadidos los puntos restantes.
- ✓ **newsizes:** Realiza la misma función que el argumento **sizes** sin embargo trabaja para el nuevo conjunto de datos añadido. Especifique nuevamente un tamaño de muestras de 300 de la forma **newsizes=300**.
- ✓ **newdata.name:** Función análoga a **data.name** para el conjunto de datos añadido. Para este gráfico especifíquelo como **newdata.name='datos ciclo PHVA'**.

De esta manera, en la *Figura 4* podrá apreciar en los recuadros de color rojo los resultados de los argumentos modificados a la carta de control resultante:

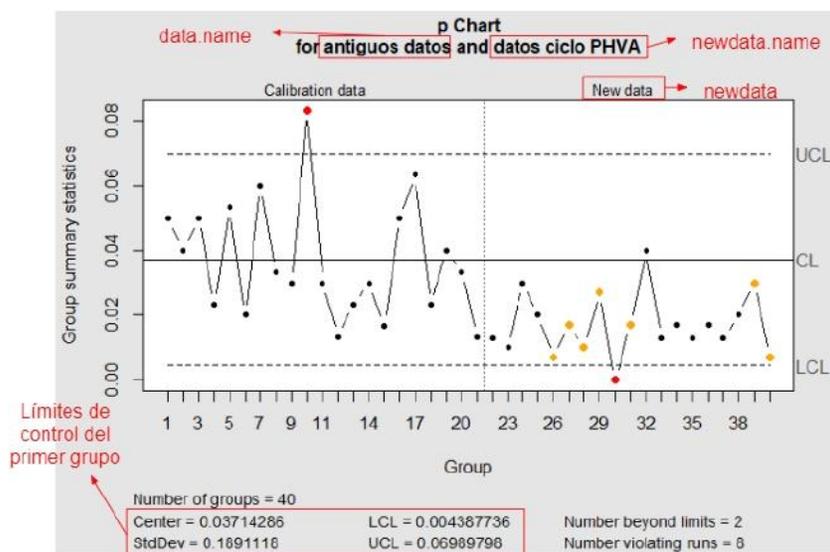


Figura 4: Carta de control con los datos añadidos del ciclo *PHVA*.

## Revisión de los límites de control del nuevo conjunto de datos.

Una vez se han realizado las respectivas oportunidades de mejora se puede apreciar un avance significativo en los nuevos 19 puntos, teniendo un sólo punto que sobrepasa el límite central. Por otra parte, los puntos que se encuentran en amarillo catalogados como *corridos o violations* no están indicando este inconveniente *per se*, si no que simplemente están haciendo parte del argumento `run.length` de la función `qcc.options`. Recuerde que este argumento lo que indica es un número de puntos consecutivos que se deben contar ya sea arriba o debajo de la línea central para a partir de este determinar los siguientes si continúan este comportamiento como corridas y cuyo valor por defecto es 7 (ejecutar la línea `qcc.options()` para ver todos los parámetros de esta función en la consola). De esta manera se aprecian que para cada primer punto amarillo de los dos grupos le anteceden seis debajo de la línea central.

En cuanto a los límites de control, al ser los valores de los argumentos `sizes` y `newsizes` la misma cifra (300) sólomente se apreciarán los límites de control de la primera carta de control generada. Esto hace que se evidencie un punto fuera de control en los nuevos datos añadidos a pesar de que este sea de valor cero.

Para poder evidenciar los límites de control de estos nuevos puntos, simplemente deberá generar una carta de control sólo para el conjunto de datos añadido de la forma que se evidencia en el siguiente comando:

```
qcc(nuevo,type = 'p', sizes = 300, title = 'Carta p para datos ciclo PHVA')
```

Con lo que haciendo uso de la función `qcc` deberá colocar como `data` el nombre del vector (“nuevo”); seleccionando el `type` “p”; dando el tamaño de 300 para `sizes` y colocando como título del gráfico en el argumento `title` “Carta p para datos ciclo PHVA”. De esta manera en la *Figura 5* se podrán apreciar los límites de control para este conjunto de datos en el nuevo gráfico de control.

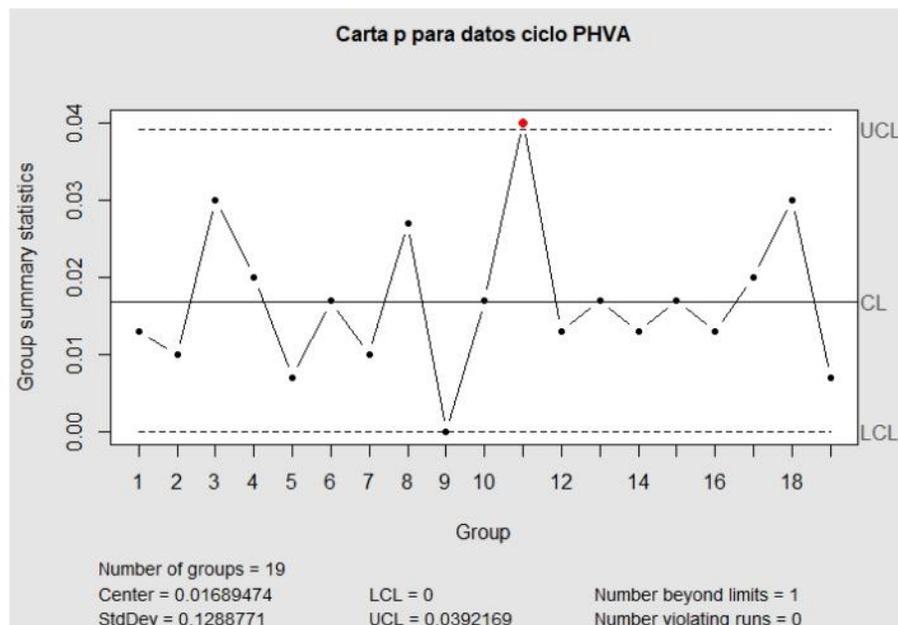


Figura 5: Carta *p* sólo para los 19 datos obtenidos después de realizar el ciclo *PHVA*.

### Carta $p$ con límites variables.

En *Rstudio* la carta de control  $p$  requiere del número de artículos defectuosos y del tamaño de la muestra para poder realizarla. En el caso de que estos sean variables se deberá crear un vector que corresponda a cada tamaño de muestra o lote  $n_i$  y establecerlo en el argumento **sizes** con el nombre que le haya asignado, o bien en el mismo argumento establecer estos valores directamente como un vector.

Para este caso se replicará el código correspondiente al utilizado para elaborar la carta de la *Figura 1* recuerde que los comando utilizados fueron:

```
library("qcc")
attach(Datos_carta_p)
Datos_carta_p
qcc(Artículos_defectuosos,type = 'p',sizes = 300)
```

De acuerdo al párrafo inicial de este inciso, se deberá realizar la modificación del argumento **sizes** cuyo valor inicial era 300 por los respectivos valores de los tamaños de las muestras los cuales están en la *Tabla 1* en la columna "Tamaño de lote". Como hizo uso de la función **attach** puede escribir el nombre directamente y se desplegará un menú en el que podrá seleccionarlo. También puede cambiar el título del gráfico en el argumento **title** por "Carta  $p$  con límites variables". De esta manera el código ahora tiene la siguiente estructura:

```
qcc(Artículos_defectuosos,type = 'p',sizes = Tamaño_de_lote )
```

En la *Figura 6* puede apreciarse que los límites de control superior e inferior se presentan de forma desprolija, evidenciándose notablemente estas variaciones en los puntos en donde el tamaño del lote es diferente a 300.

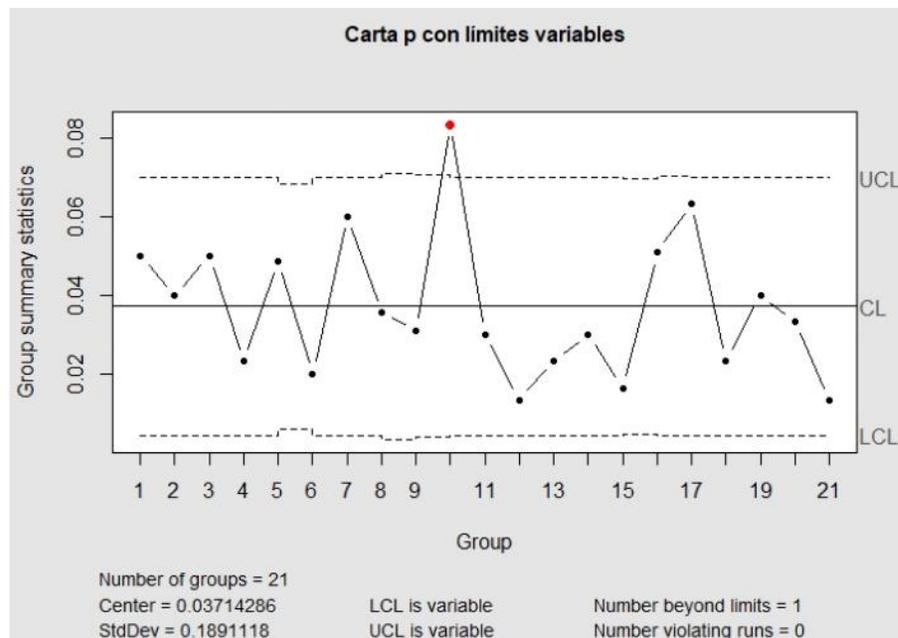


Figura 6: Carta  $p$  con límites de control variables.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

A partir de la información suministrada resolver los siguientes puntos:

*En un proceso se produce por lotes y éstos se prueban a 100 %. Se lleva un registro de la proporción de artículos defectuosos. Los datos de los últimos 25 lotes se muestran en la Tabla 2.*

1. Obtenga una carta  $p$  usando el tamaño de muestreo (lote) promedio e interprete.
2. ¿Cómo explicaría los límites de control que obtuvo alguien que no tiene conocimientos profundos de estadística?
3. Obtenga una carta  $p$  con límites de control variables e interprete.
4. ¿Que límites de control usaría para analizar datos futuros mediante las cartas  $p$ ?

*Tabla 2 correspondiente a artículos defectuosos en 24 lotes.*

Lote	Tamaño	Artículos defectuosos
1	200	21
2	200	20
3	200	27
4	200	33
5	200	22
6	200	40
7	180	27
8	180	23
9	180	20
10	200	26
11	200	28
12	200	21
13	200	23
14	200	21

Lote	Tamaño	Artículos defectuosos
15	200	25
16	200	29
17	200	20
18	220	28
19	220	18
20	220	24
21	200	13
22	200	23
23	200	12
24	200	19
25	200	26

## 6. Referencias

- [1] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010

# PRACTICA 10: CARTA DE CONTROL PARA NÚMERO DE ARTÍCULOS DEFECTUOSOS NP

## 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control para número de artículos defectuosos **np** mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Identificación de los contextos en los que requiera hacer uso de entre una carta de control *p* y una carta *np*
- ✓ Interpretación y análisis de la información del gráfico resultante.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

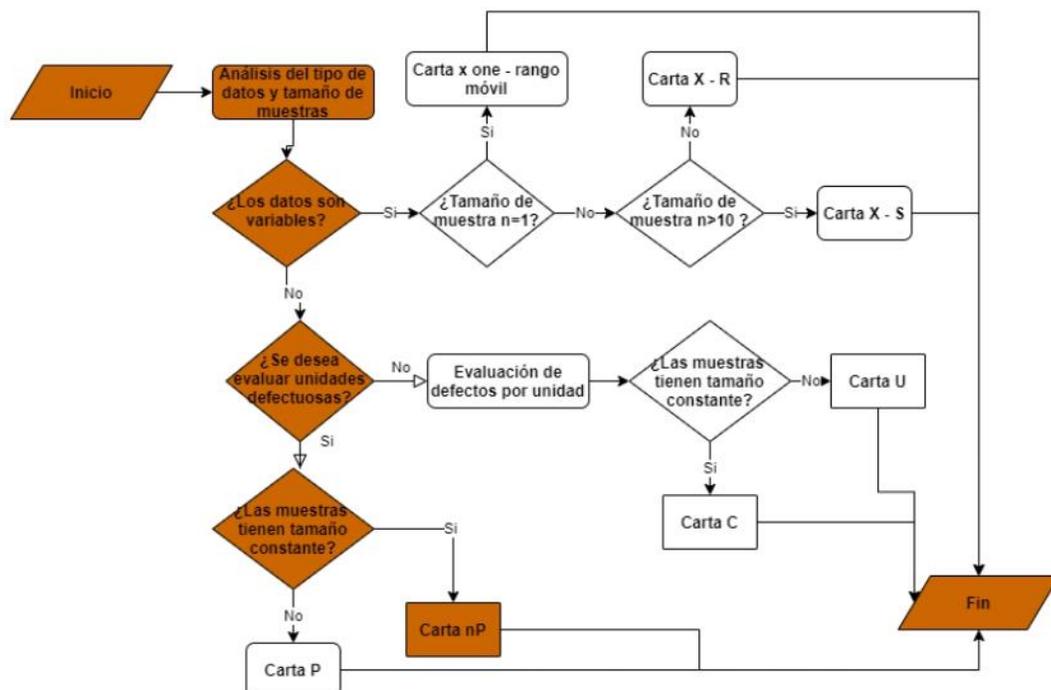


Figura 1: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta np.

### 3.1 Definición.

La carta de control para el número de artículos defectuosos es una variación de la carta  $p$  en la cual se gráficán en vez de proporciones, el número de defectos. Esto es útil en casos en los que el tamaño de la muestra  $n_i$  sea constante [1]. Partiendo de la base de que, se sigue una distribución binomial, los límites de control para esta carta están dados por:

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1 - \bar{p})} \quad (1)$$

$$CL = n\bar{p} \quad (2)$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1 - \bar{p})} \quad (3)$$

Estas cartas de control permiten visualizar la realidad del proceso en base en la forma de muestreo de los datos, con lo que es importante el hecho de la respectiva identificación de causas asignables que estén presentes en el proceso. Otro factor importante a tener en cuenta es el caso en el que se presente un valor negativo en el límite de control inferior con lo cual, habrá que revisar con detalle errores en la elaboración de estos parámetros o en su defecto, deberá optar por establecer que este límite tenga un valor de cero ya que, es lógicamente imposible que existan unidades defectuosas negativas.

#### Ventajas, desventajas y usos.

La carta  $np$  resulta de fácil lectura para aquellas personas con un conocimiento básico de estadística ya que, simplemente se están trabajando con valores enteros. Por otra parte, al ser de tamaño de muestra constante, la carta resulta ser prácticamente la misma a la carta  $p$  con lo que para convertir una carta de proporción en una de unidades defectuosas basta dividir la escala de la carta en el tamaño de la muestras. Por lo tanto, se puede decir que el criterio para escoger una carta de otra depende de la disposición del análisis deseado entre proporción o unidades, esto partiendo de la base que se menciona anteriormente de que se trate de tamaños de muestra constantes.

Por otra parte, este gráfico presenta la desventaja de que, si bien se puede trabajar esta carta con tamaños de muestras variables, su lectura presenta ser en demasía una inexorable complejidad [2]. Más aún al requerir un tipo de carta estandarizada, por lo cual, en estos casos se recomienda mucho más el uso de la carta  $p$ .

## 4. Montaje y procedimiento

**Docente:** Indique al estudiante que abra el software *Rstudio* y que elabore una carta para proporción de artículos defectuosos o *Carta np* a partir de los datos del ejercicio planteado. A continuación, realice la explicación paso a paso de las funciones y del código correspondiente a la elaboración de los gráficos.

**Estudiante:** Siga las instrucciones del docente detalladamente en pro de la comprensión a la hora de elaborar la carta  $np$ . Para lograr este objetivo, se partirá del siguiente caso:

## Planteamiento del problema.

Del análisis de los datos de inspecciones y pruebas finales de un producto ensamblado, se ha detectado, a través de una estratificación y un análisis de Pareto, que la causa principal por la que los artículos salen defectuosos está relacionada con los problemas en un componente en particular: el componente W. Se decide, entonces, analizar más de cerca el proceso que produce tal componente. Para ello, de cada lote de componentes W se inspecciona una muestra de  $n = 120$ . [1] Los datos obtenidos en 20 lotes consecutivos se muestran en la Tabla 1.

Tabla 1: Número de artículos defectuosos para un componente w.

Lote	Artículos_defectuosos
1	9
2	6
3	10
4	8
5	5
6	5
7	14
8	12
9	9
10	8
11	10
12	20
13	12
14	10
15	10
16	0
17	13
18	5
19	6
20	11

! Recuerde que puede hacer la importación en la ventana superior derecha del software, dando click en el botón "Import Dataset" y luego dando click en "From Excel.."

## Elaboración del código.

El código correspondiente para la elaboración de esta carta de control se evidencia y explica paso a paso a continuación:

```
library("qcc")
attach(Datos_carta_nP)
head(Datos_carta_nP)
qcc(Artículos_defectuosos,type="np",sizes=120,
title='Carta nP para número de artículos defectuoso de un producto ensamblado')
```

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escribálas en *Rstudio*.

1. Haga la puesta en escena de la librería *qcc*.
2. Utilice `attach()` para poder llamar directamente la columna con el número de artículos defectuosos
3. Utilice la función `head()` para poder visualizar los nombres de las columnas y sus seis primeros datos.
4. Mediante la función `qcc` realice la carta de control respectiva, colocando en `data` el nombre de la segunda columna; para `type` el tipo “np” y finalmente en `sizes` dictamine un tamaño de muestra de 120.
5. Finalmente, haciendo uso del argumento `title` especifique de nombre al gráfico de la forma `title='Carta nP para número de artículos defectuoso de un producto ensamblado'`

De esta manera, la carta resultante tiene la siguiente apariencia:

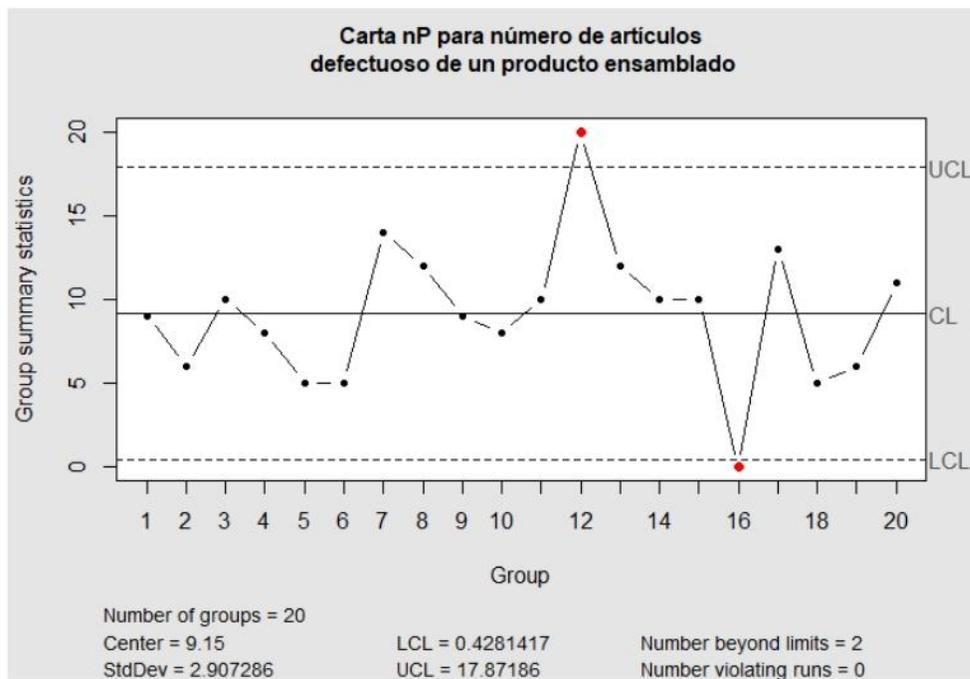


Figura 2: Carta *np* correspondiente al número de defectos de un producto ensamblado.

### Otros argumentos de la función `qcc`

- ✓ **labels:** Un vector de caracteres o bien numérico, en el cual se establecen las etiquetas que deben tener los valores de las muestras en el eje *X*. Por defecto *Rstudio* coloca estos nombres como valores. Tiene que tener presente que debe colocar estrictamente nombre a cada dato que fue especificado, de lo contrario saldrá un error en la consola del siguiente tipo:

```
Error in dimnames(x) <- dn :
la longitud de 'dimnames' [1] no es igual a la extensión del arreglo
```

- ✓ **rules:** Conjunto de reglas que se aplican a la carta de control, es decir, unos parámetros los cuales realizarán todos los cálculos respectivos. Por defecto este parámetro utiliza la función `shewhart.rules`. La recomendación es dejar este argumento por defecto.
- ✓ **add.stats:** Argumento lógico (*TRUE* o *FALSE*) cuya función es la de decidir si la información estadística que se encuentra en la parte inferior del gráfico debe ser mostrada.
- ✓ **chart.all:** Argumento lógico que indica si se deben plotear ambos conjuntos declarados en **data** y **newdata**. Si se declara *FALSE*, sólo se graficará el primer conjunto de datos (los del argumento **data**).
- ✓ **label.limits:** Vector que permite especificar las etiquetas que deben tener los límites de control y superior respectivamente. Puede realizar la puesta en escena de este argumento haciendo uso de la forma `label.limits=c('LCL','UCL')` que hacen alusión a las siglas en ingles de los límites de control superior e inferior.
- ✓ **digits:** Único valor cuyo función es indicar el número de cifras significativas que deben mostrarse en informativo de la parte inferior del gráfico (más específicamente para los valores de *Center*, *stdDev* y los límites de control).
- ✓ **restore.par:** Este argumento lógico tiene la función de restablecer los parámetros de la carta de control a su estado por defecto si hizo uso de la función **par**. Si desea agregar más puntos al gráfico se aconseja dictaminar este argumento como falso.

! La función **par** cuenta con una amplia variedad de parámetros que permiten realizar modificaciones a los gráficos realizados en *Rstudio*. Si desea conocer más sobre esta función puede consultar el manual *R: A language and Environment for Statistical Computing* o bien pedir indicación a su docente.

**Función: Shewhart rules.**

Esta función permite de una manera sencilla llamar en la consola aquellos puntos que se encuentra fuera de control en el gráfico y cuales son los se encuentran como corridas. Su forma general corresponde a:

```
shewhart.rules(object, limits = object$limits,
run.length = qcc.options('run.length'))
```

Para que esta función dé resultado deberá asignar un nombre a la línea utilizada para hacer el gráfico de control de tal forma que quede con la siguiente estructura:

```
test<-qcc(Artículos_defectuosos,type="np",sizes=120,
title="Carta nP para número de artículos defectuoso de un producto ensamblado")
```

De tal que "test" será catalogado como un **objeto de clase qcc**. De esta manera, la forma de los argumentos de la función `shewhart.rules` son:

- ✓ **object:** Objeto de clase *qcc*, en este caso será el objeto que se ha dictaminado para realizar el gráfico de control

- ✓ **limits:** Límites de control. La forma de uso es llamar al objeto de clase *qcc* del argumento anterior, agregar el signo \$ y seleccionar en la lista que se despliega la opción *limits*.
- ✓ **run.length:** Es el mismo argumento que se utiliza en la función *qcc.options*. La diferencia radica en que si se especifica un valor en este apartado no quedará por defecto para toda la librería.

Con lo cual, el código a ejecutar para esta ejercicio corresponde a:

```
shewhart.rules(test,test$limits,run.length = 7)
```

Cuando ejecuta la función, en la consola aparecerán los valores de las funciones **beyond.limits** y **violating.runs**, las cuales respectivamente son las que se encargan de mostrar los puntos fuera de control y los puntos de corrida. Puede llamar individualmente a cada función utilizando los mismos argumentos mencionados anteriormente de la siguiente forma:

- ✓ Para la función **beyond.limits** utilizará los argumentos de la función *shewhart.rules* **object** y **limits** de la misma forma utilizada anteriormente. Use la línea de tipo `beyond.limits(test,test$limits)` para su respectivo uso.
- ✓ Por otra parte, si desea utilizar la función **violating.runs** deberá utilizar la línea de la forma `violating.runs(test, run.length = 7)` en la cual solamente se están utilizando los argumentos **object** y **run.length**. Recuerde que se asigna el valor de siete debido a que es el valor por defecto de la cantidad de puntos precedentes antes de empezar a contar puntos como corridas de forma consecutiva ya sea arriba o abajo de la línea central.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuesta deben estar relacionadas con lo observado durante la práctica.

De acuerdo al problema planteado responda las siguientes preguntas:

*Para analizar la estabilidad de la cantidad de artículos defectuosos en un proceso de producción y tratar de mejorarlo, se toma una muestra de 120 piezas cada cuatro horas, mediante el método de intervalo. Los datos obtenidos durante seis días se muestran en la Tabla 3:*

Tabla 3: Datos de 20 muestras de artículos defectuosos.

Muestra	Artículos defectuosos	Muestra	Artículos defectuosos
1	11	11	8
2	10	12	7
3	7	13	9
4	10	14	6
5	4	15	6
6	12	16	11
7	8	17	9
8	5	18	7
9	14	19	6
10	12	20	10

1. Para analizar estos datos, ¿cuál carta ( $p$  o  $np$ ) recomendaría?
2. Mediante una carta  $p$  analice los datos y comente los resultados.
3. ¿Cómo explicaría los límites de control que obtuvo?
4. De acuerdo con los costos de producción, el nivel de artículos defectuosos máximo tolerable es de 10%. Por lo tanto, se sugiere que el límite de control superior de la carta  $p$  sea 0.10. ¿Es correcta esta sugerencia?
5. Analice los datos con una carta  $np$  e interprete.

## 6. Referencias

- [1] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010
- [2] Sanchez, Ismael. *Apuntes para la asignatura Métodos Estadísticos para la Mejora de la Calidad*. Universidad Carlos III de Madrid, sin fecha

# PRACTICA 11: CARTA DE CONTROL PARA NÚMERO DE DEFECTOS POR SUBGRUPO C

## 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control para número de defectos por subgrupo  $c$  mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Reconocer los parámetros a modificar para el desarrollo pertinente de estos gráficos de control.
- ✓ Interpretación y análisis de la información del gráfico resultante.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

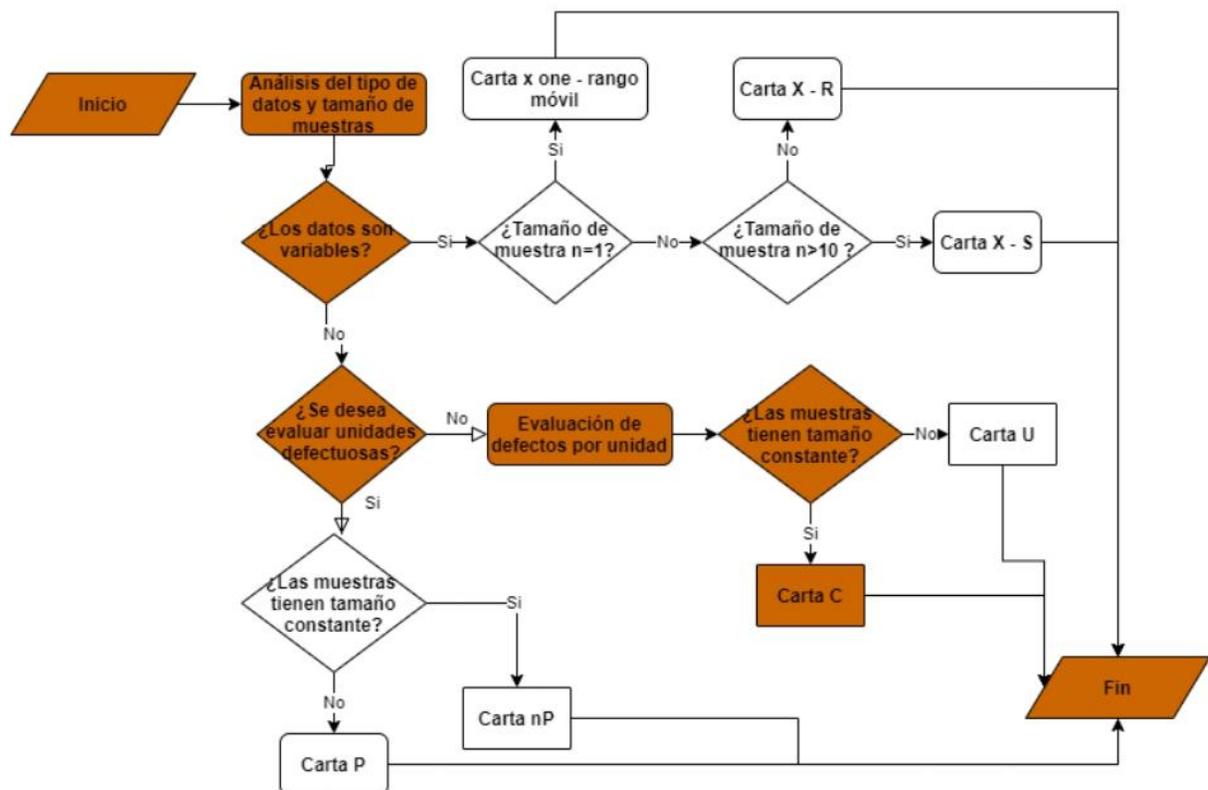


Figura 1: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta  $c$ .

### 3.1 Definición.

Un artículo *no conforme* es una unidad que que cumple con una o varias especificaciones dadas para el producto. Cada punto en el que no se cumplen las especificaciones se denomina como *defecto o disconformidad* sin embargo, dependiendo de la magnitud y naturaleza de este se puede tomar la decisión de no clasificarlo como *no conforme* debido a que pueden ser pequeñas imperfecciones que no implican una ineficiencia en el uso del producto. Caso contrario sucede con aquellas unidades que poseen defectos considerables que acarrear dificultades para el cliente.[1]

Existen ocasiones en las que es mucho más conveniente trabajar con el número de defectos por unidad que simplemente con la proporción no conforme con lo cual, a la hora de medir esta cantidad de defectos en los productos entra en escena el uso de la carta de control para número de defectos por subgrupo o más conocida como *carta c*.

#### Cálculo de límites de control.

La carta de control *c* tiene como objetivo analizar el número de defectos por subgrupo cuando el número del subgrupo se mantiene constante[2]. Partiendo de la base de una distribución de *Poisson* el cálculo de la media y la desviación corresponden a:

$$\mu_{c_i} = \bar{c} = \frac{\text{Total de defectos}}{\text{Total de subgrupos}} \quad (1)$$

$$\sigma_{c_i} = \sqrt{\bar{c}} \quad (2)$$

De esta forma los límites de control respectivos para esta carta de control son:

$$UCL = \bar{c} + 3\sqrt{\bar{c}} \quad (3)$$

$$CL = \bar{c} \quad (4)$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} \quad (5)$$

Se recalca con suma importancia de que el propósito vital de la carta *c* es el de analizar el número de defectos por subgrupo o bien por unidad lo cual hace referencia a que puede ser un artículo o un lote, una medida de longitud o de tiempo, una medida de área o volumen **de tamaño constante**. Cuando la dimensión del subgrupo es variable se acude al uso de la *carta u* la cual se estudiará con más detalle en la siguiente práctica.

#### Aplicaciones

Por otra parte en cuanto respecta a los casos de uso, esta carta de control aplica en puntos de inspección en los cuales se busca localizar uno o más tipos de defectos relativamente menores o tolerables, de tal forma que, aunque se encuentren estos errores, no se clasifique la unidad como *no conforme* si no sobrepasa el umbral de defectos establecidos por la organización. Algunos ejemplos de su uso aplican para:

- ✓ Número de quejas.
- ✓ Número de errores.
- ✓ Cantidad de paros o detenciones.

Se recomienda que, de ser posible se elijan tamaños de subgrupo de forma tal que el número promedio de defectos por subgrupo (línea central) sea mayor que nueve. Por otra parte, si en cada muestra se esperan cero o muy pocos defectos, mucho menos que nueve, usualmente la carta no es eficaz.[2] En esos casos, se aconseja incrementar el tamaño del *sample* o buscar otras alternativas.

#### 4. Montaje y procedimiento

**Docente:** Indique al estudiante que abra el software *Rstudio* y que elabore una carta para proporción de defectos por subgrupo o *Carta c* a partir de los datos del ejercicio planteado con lo cual deberá registrar estos datos en un archivo en Excel e importarlos a *Rstudio*. A continuación, realice la explicación paso a paso de las funciones y del código correspondiente a la elaboración de los gráficos.

**Estudiante:** Siga las instrucciones del docente detalladamente en pro de la comprensión a la hora de elaborar la carta *c*. Para lograr este objetivo, se partirá del siguiente caso:

! Recuerde que puede hacer la importación en la ventana superior derecha del software, dando click en el botón “Import Dataset” y luego dando click en “From Excel...”

*En una empresa que elabora productos agroquímicos se tiene el problema de intoxicación de los trabajadores debido al contacto con agentes tóxicos. Para evaluar el número de obreros intoxicados por mes en los últimos dos años, se recurre a los registros de la enfermería de la empresa. Los datos obtenidos se muestran en la Tabla 1:*

Tabla 1: Datos de personas intoxicadas por mes.

Mes	Intoxicados	Mes	Intoxicados	Mes	Intoxicados
1	6	7	3	13	5
2	5	8	7	14	4
3	4	9	5	15	7
4	4	10	7	16	2
5	1	11	5	17	4
6	3	12	12	18	2

#### Elaboración del código

El correspondiente paso a paso del código respectivo se evidencia a continuación:

1. Haga la puesta en escena de la librería *qcc* con la función `library`.
2. Utilice la función `attach` para llamar directamente las columnas por su respectivo título colocando el nombre del conjunto de datos que ha importado a *Rstudio*.
3. Con la función `head` puede obtener una breve visualización de los datos en la consola.
4. Mediante la función `qcc` realice la carta de control *c*. Recuerde que la forma general de esta función corresponde a la siguiente estructura:

! Cuando se habla de la forma general de una función, esto sólo indica la forma correcta de su uso y el tipo de argumentos que esta utiliza (lógico, numérico, vector, entre otros). **No se debe ejecutar esta línea.**

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

Deberá colocar en el argumento **data** el nombre de la segunda columna (sin comillas); en **type** seleccione el tipo “c”; finalmente para el argumento **title** dictamine “Carta c para número de intoxicados por mes”.

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

En este orden de ideas, la línea de comandos presenta la siguiente estructura:

```
library('qcc')
attach(Ejemplo_carta_c_Hoja_1)
head(Ejemplo_carta_c_Hoja_1)
qcc(Intoxicados,type = 'c',title = 'Carta c para número de intoxicados por mes')
```

Para esta carta de control no es necesario asignar un tamaño de subgrupo en el argumento **sizes** ya que esta trabaja con tamaños constantes. De esta manera el gráfico resultante presenta los siguientes datos:

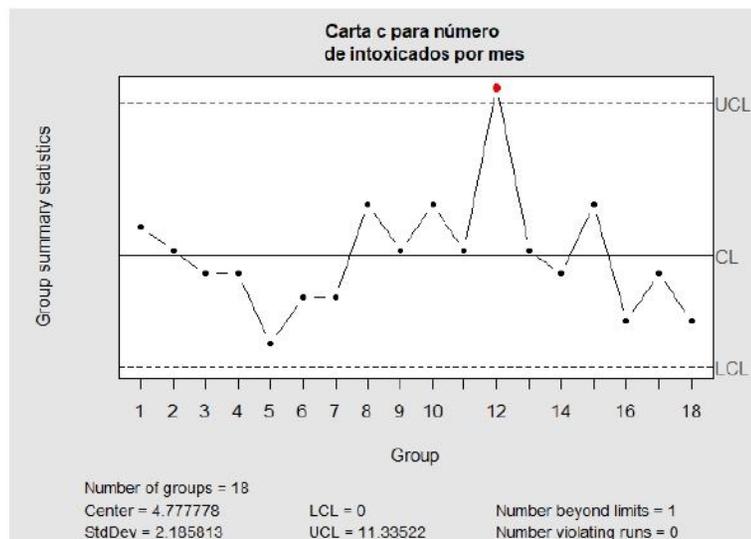


Figura 2: Grafico de control c para número de intoxicados por mes

### Funciones stats. sd. limits.

Estas funciones son las que realizan los cálculos correspondientes a las cartas. Para esto deberemos asignar un objeto de clase `qcc` a la línea de código correspondiente a la elaboración de carta de control, de esta manera el nuevo comando a elaborar es:

```
comodin<-qcc(Intoxicados,tye = 'c',title = 'Carta c para número de intoxicados por mes')
```

Esto para poder acceder a cada uno de los atributos evidenciados en la consola de manera individual. De tal manera el código deberá tener la estructura:

! Recuerde que puede llamar los atributos del objeto qcc creado colocando el nombre de este y añadiendo un signo pesos (\$) con lo que se desplegará una lista con el atributo que desea utilizar.

```
stats.c(Intoxicados,sizes = comodin$sizes)
sd.c(Intoxicados,std.dev=NA,sizes=comodin$sizes)
limits.c(center = 4.777778,std.dev = 2.185813,sizes = comodin$sizes,conf = 0.99)
```

La estructura y los argumentos de cada una de estas funciones serán explicados a continuación

- ✓ **stats.:** Esta función se encarga de realizar los estadísticos para puntos respectivos según la carta requerida (puntos para carta de rangos, medias, proporción de unidades defectuosas, etc).

Tabla 2: principales tipos de cartas para las funciones stats. , sd. y limits.

Función	Tipo de carta
stats. sd. y limits.	"xbar"
	"R"
	"S"
	"xbar"
	"xbar.one"
	"p"
	"np"
	"c"
	"u"
	"g"

Sus respectivos argumentos son

- **data:** El conjunto de datos
  - **sizes:** El tamaño de las muestras.
- ✓ **sd.:** Calcula la desviación estandar de la carta solicitada. También hace uso del argumento **data** y cumple la misma función descrita, además utiliza el siguiente argumento:
    - **std.dev:** Valor numérico o un método para calcular la desviación estándar (UWAVE-R,UWAVE-SD, entre otros), para las cartas de control por atributos se dictamina como NA.
  - ✓ **limits:** Calcula los límites de control. Esta se compone de los siguientes argumentos:
    - **center:** Valor de la línea central (este se puede obtener de los atributos de la carta de control, una vez ejecute el comando de la carta esta información saldrá en la consola, o bien, si le ha asignado un nombre al comando para plotear la carta de control, lo puede visualizar escribiendo el nombre asignado y añadiendo el signo \$, con lo que desplegará la lista de atributos).
    - **std.dev:** Valor que corresponde a la desviación estándar. Aquí se debe colocar el valor resultante de la función **sd.**
    - **sizes:** Valor correspondiente al tamaño de la muestra.
    - **conf:** Valor numérico el cual especifica el número de desviaciones estándar si se establece un valor mayor a uno. Un valor entre 0 y 1 dictamina el nivel de confianza.

En la *Figura 3* se evidencia el resultado en la consola de la puesta en escena de cada uno de las funciones:

```

> stats.c(Intoxicados,sizes = comodin$sizes)
$statistics
[1] 6 5 4 4 1 3 3 7 5 7 5 12 5 4 7 2 4 2

$center
[1] 4.777778

> sd.c(Intoxicados,std.dev=NA,size=comodin$sizes)
[1] 2.185813
> limits.c(center = 4.777778,std.dev = 2.185813,sizes = comodin$sizes,conf = 0.9
9)
LCL UCL
0 11

```

Figura 3: Puesta en escena de la funcione `stats`, `sd` y `limits`

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

A partir del problema planteado resolver las siguientes preguntas:

Con el propósito de analizar la posibilidad de eliminar los estándares de trabajo en un sector de una fábrica, se decide analizar el número de cierto tipo de operaciones que realiza cada trabajador por día y semana. En la tabla 2 se muestran los resultados obtenidos en una semana para 14 trabajadores.

Tabla 2: datos para trabajadores por semana

Trabajador	Operaciones
1	295
2	306
3	292
4	297
5	294
6	343
7	285
8	240
9	329
10	305
11	277
12	260
13	337
14	320

1. Calcule los límites de control para una carta  $c$  en base al número de operaciones por trabajador e interprete los límites que obtenga.
2. Investigue, mediante una carta  $c$  si algún trabajador está fuera del sistema, es decir, si alguno se encuentra realizando operaciones por fuera de los límites de control establecidos. En caso de estarlo, ¿qué recomendaría que se hiciera con dicho trabajador?

3. Analice esos mismos datos mediante una carta  $X - One$  para mediciones individuales.
4. ¿Cuál de las dos cartas parece ser más apropiada?

## 6. Referencias

- [1] D. V. Montgomery, Douglas C y Verbeek. *Control estadístico de la calidad*. Limusa Wiley, 2004.
- [2] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010

# PRACTICA 12: CARTA DE CONTROL PARA NÚMERO PROMEDIO DE DEFECTOS POR UNIDAD U

## 1. Competencias a desarrollar

- ✓ Desarrollar los conocimientos necesarios para poder realizar correctamente cartas de control para número promedio de defectos por unidad  $u$  mediante el paquete *Quality Control Charts (qcc)*.
- ✓ Interpretación y análisis de la información del gráfico resultante.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

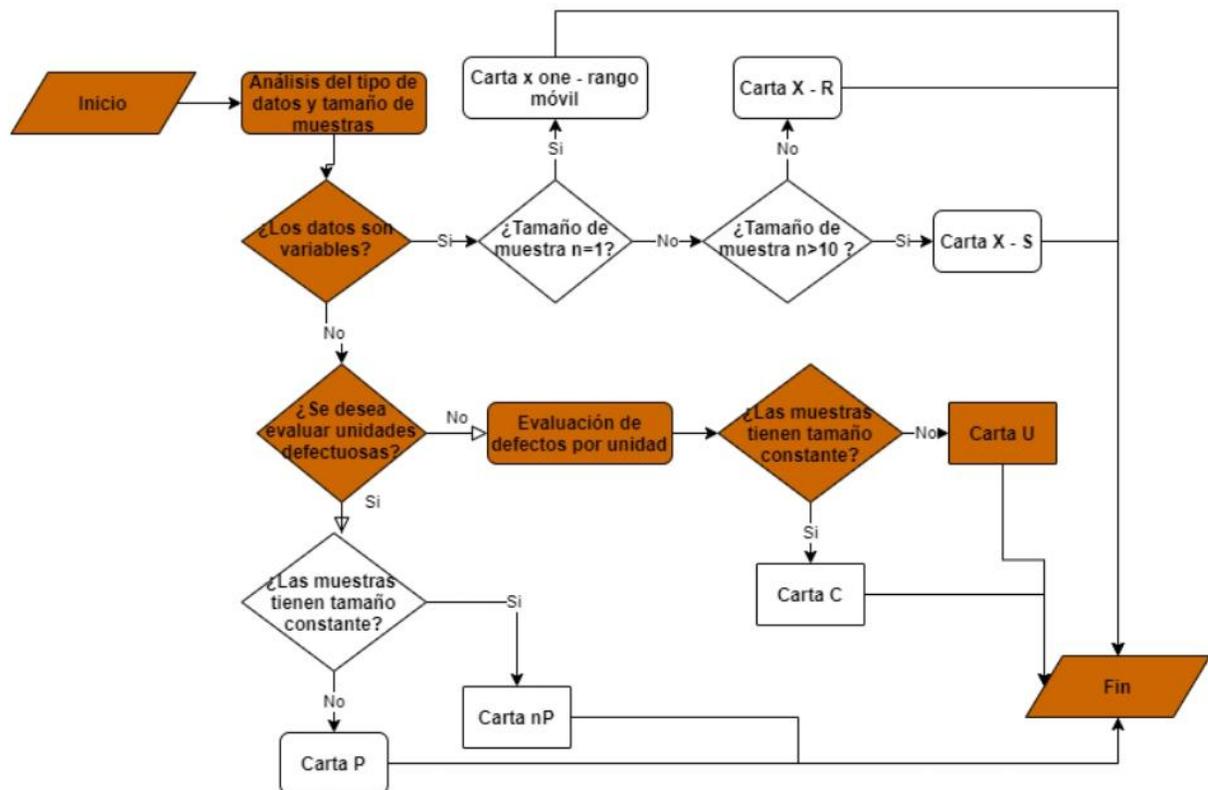


Figura 1: Diagrama de flujo de análisis en elección de carta de control indicando en color naranja los acontecimientos para escoger la carta  $u$ .

### 3.1 Definición.

La carta de control para medición de defectos promedio por unidad o más conocida como **carta u** se utiliza cuando el tamaño de la muestra o subgrupo no son constantes.[1] Siguiendo una distribución de *Poisson* se procede para calcular la media y la desviación de la forma:

$$\mu_{u_i} = \bar{u} = \frac{\text{Total de defectos}}{\text{Total de artículos inspeccionados}} \quad (1)$$

$$\sigma_{u_i} = \sqrt{\frac{\bar{u}}{n}} \quad (2)$$

Con lo cual se tiene que los límites de control respectivos para esta carta son:

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} \quad (3)$$

$$UCL = \bar{u} \quad (4)$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} \quad (5)$$

#### Usos y aplicaciones.

En lo que respecta a los usos y aplicaciones de la carta *u* aplica igual que la carta *c*, pero aquí se prefiere analizar el número promedio de defectos por artículo o unidad en lugar del número de defectos por subgrupo. Algunos ejemplos son:

- ✓ Número de no conformidades por lote o subgrupo de materia prima recibida donde el tamaño del lote varía.
- ✓ Cantidad de nuevas infecciones en un hospital por día.
- ✓ Cantidad de accidentes por camión de reparto por día

Al igual que ocurre con su gráfico análogo la carta *c* si el tamaño de la muestra *n* es mucho menor que el número recomendado (9), la carta *u* no suele ser útil. En esos casos, hay que buscar incrementar *n* o utilizar otra carta de control.[1] Cabe recalcar que, esta carta es más efectiva cuando se toma para analizar tamaños de subgrupos con variables, sin embargo también es posible utilizarla con muestras constantes calculando el número promedio de subgrupo, lo cual es eficaz cuando en el total de estos se presenta una variación menor al 20% aproximadamente.

Un aspecto importante es que esta carta de control presenta de manera análoga los mismos inconvenientes con su hermana la carta *p* en lo que respecta cuando el tamaño de los subgrupos no es constante. Para ese caso también se opta por la opción ya mencionada de una proporción promedio o bien realizar una **carta u estandarizada** la cual corresponde en este caso a la ecuación 4:

$$Z_i = \frac{u_i - \bar{u}}{\sqrt{\frac{\bar{u}}{n_i}}} \quad (4)$$

## 4. Montaje y procedimiento

### Carta $u$ con límites constantes.

**Docente:** Indique al estudiante abrir el software *Rstudio* y que elabore una carta para proporción de artículos defectuosos o *carta u* a partir de los datos del ejercicio planteado. A continuación, realice la explicación paso a paso de las funciones y del código correspondiente a la elaboración de los gráficos.

**Estudiante:** Siga las instrucciones del docente detalladamente en pro de la comprensión a la hora de elaborar la carta  $u$ . Para lograr este objetivo, se partirá del siguiente caso:

### Planteamiento del problema.

*Un fabricante de computadoras personales desea establecer una carta de control para las disconformidades por unidad (carta  $u$ ) en la línea de ensamblaje final. El tamaño de la muestra se selecciona de cinco computadoras. En la Tabla 1 se muestran los datos del número de disconformidades en 20 muestras de 5 computadoras cada una.*

Tabla 1. Datos de defectos por unidad en 20 muestras con tamaño de subgrupo 5

Número de muestra	Tamaño de la muestra	Número total de disconformidades	Numero promedio de disconformidades por unidad
1	5	10	2.0
2	5	12	2.4
3	5	8	1.6
4	5	14	2.8
5	5	10	2.0
6	5	16	3.2
7	5	11	2.2
8	5	7	1.4
9	5	10	2.0
10	5	15	3.0
11	5	9	1.8
12	5	5	1.0
13	5	7	1.4
14	5	11	2.2
15	5	12	2.4
16	5	6	1.2
17	5	8	1.6
18	5	10	2.0
19	5	7	1.4
20	5	5	1.0

### Elaboración del código.

El paso a paso de la línea de comandos respectiva se explica a continuación:

1. Haga la puesta en escena de la librería *qcc* con la función `library`.

2. Elabore un vector con los datos de la columna 3 de la *Tabla 1* y asigne el nombre de “Número\_disconformidades”.
3. Deberá acudir nuevamente al uso de la función `qcc`. Recuerde que la forma general de esta función se establece con la siguiente estructura:

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

Deberá colocar en el argumento **data** el nombre del vector creado anteriormente; en **type** seleccione el tipo “u” para realizar el tipo de carta deseado; en el argumento **sizes** asigne un valor de 5 debido a que el tamaño para todas las muestras es de esta magnitud, y finalmente asigne el título al gráfico de “carta u para Número promedio de defectos por unidad” en el argumento **title**.

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

En este orden de ideas, el código correspondiente cuenta con la siguiente estructura:

```
library("qcc")
Numero_disconformidades<-c(10, 12, 8, 14,....,5)
qcc(Numero_disconformidades,type = "u", sizes =5,
title="carta u para Número promedio de defectos por unidad")
```

! los puntos suspensivos que evidenció en la segunda línea de código son sólo una forma de acotar el código. Para que este funcione deberá colocar todos los valores de la columna correspondiente.

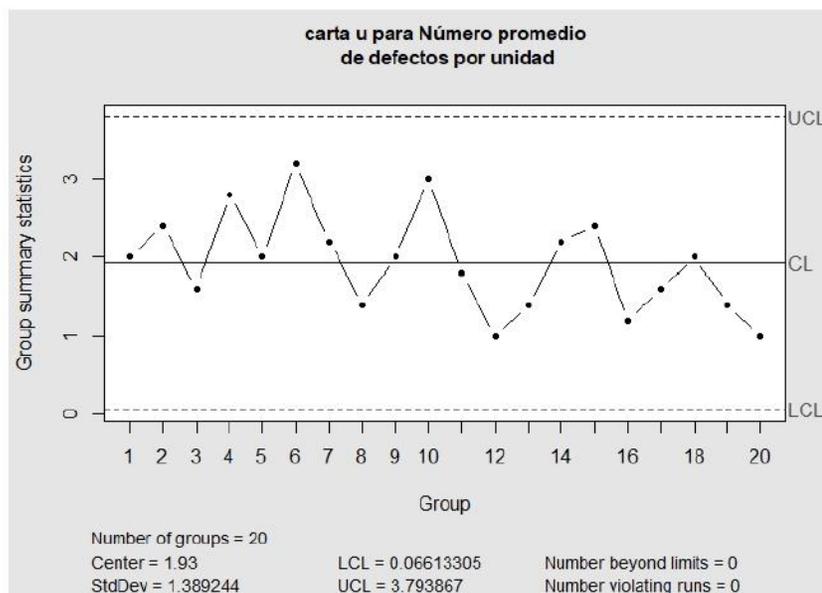


Figura 2: Carta u para defectos por unidad

De esta manera el gráfico de control resultante se evidencia en la *Figura 2*. En la carta resultante si bien el proceso aparentemente se encuentra dentro de control estadístico, se aprecian unas cifras muy elevadas en cuanto a los defectos por unidad se trata. Con lo cual, es un indicador para la administración de una necesidad imperiosa de tomar cartas en el asunto en pro de la optimización del proceso.

### Carta u con límites variables.

La diferencia principal en la elaboración de la carta *u* con límites variables radica en el argumento **sizes** de la función `qcc`, en el cual en vez de establecer una constante se dictamina un vector cuyos valores correspondan a cada tamaño de muestra o bien tomar estos datos de una columna con información de un *dataframe* elaborado. En este orden de ideas se puede disponer que, el procedimiento es análogo al realizado en prácticas anteriores con el que se utilizó para la carta *p*. A partir siguiente enunciado se ejemplifica el método de elaboración de la carta *u* con límites de control variables:

### Segundo planteamiento.

*En una planta de acabados textiles se inspecciona la tela teñida para la ocurrencia de defectos en cada 50 metros cuadrados. En la Tabla 2 se muestran los datos de diez rollos de tela. Se usarán estos datos para establecer una carta de control para las disconformidades por unidad.*

*Tabla 2: Datos para rollos de tela.*

Número de rollo	Número de metros cuadrados	Número total de disconformidades	Número de unidades de inspección en el rollo (n)	Número de disconformidades por unidad de inspección
1	500	14	10.0	1.40
2	400	12	8.0	1.50
3	650	20	13.0	1.54
4	500	11	10.0	1.10
5	475	7	9.5	0.74
6	500	10	10.0	1.00
7	600	21	12.0	1.75
8	525	16	10.5	1.52
9	600	19	12.0	1.58
10	625	23	12.5	1.84

### Elaboración del segundo código.

El código correspondiente para este fin se explica en los siguientes pasos:

1. Elabore dos vectores con los datos de las columnas 3 y 4 respectivamente. Nombre a cada uno como “disconformidades\_rollo” y “unidades\_inspección”.
2. Puede realizar un breve chequeo del resultado en la consola ejecutando el nombre de estos en el área de edición de scripts.
3. Asigne el nombre de “rollos” al comando correspondiente con la función `qcc`. Recuerde nuevamente que la forma general de esta función corresponde a:

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE)
```

En este inciso deberá declarar en los siguientes argumentos:

- ✓ Para **data** el nombre del primer vector.
- ✓ En **type** escriba nuevamente el tipo “u”.
- ✓ Para el argumento **sizes** hará uso de los datos del vector “unidades\_inspección”. Con esto *Rstudio* realizará los respectivos cálculos de los puntos respectivos y de los límites de control.
- ✓ Finalmente, especifique en el argumento **title** el nombre de “Carta u con límites variables para rollos de tela.”

De esta manera, el código presenta la siguiente estructura:

```
disconformidades_rollo<-c(14,12,20,11...,23)
unidades_inspección<-c(10.0, 8.0, 13.0, 10.0, ...,12.5)

disconformidades_rollo
unidades_inspección

rollos<-qcc(disconformidades_rollo,type = ‘‘u’’, sizes = unidades_inspección,
title=‘‘Carta u con límites variables para rollos de tela’’)
```

Una vez realizado esto, la *figura 3* muestra la representación de la carta de control con los límites variables:

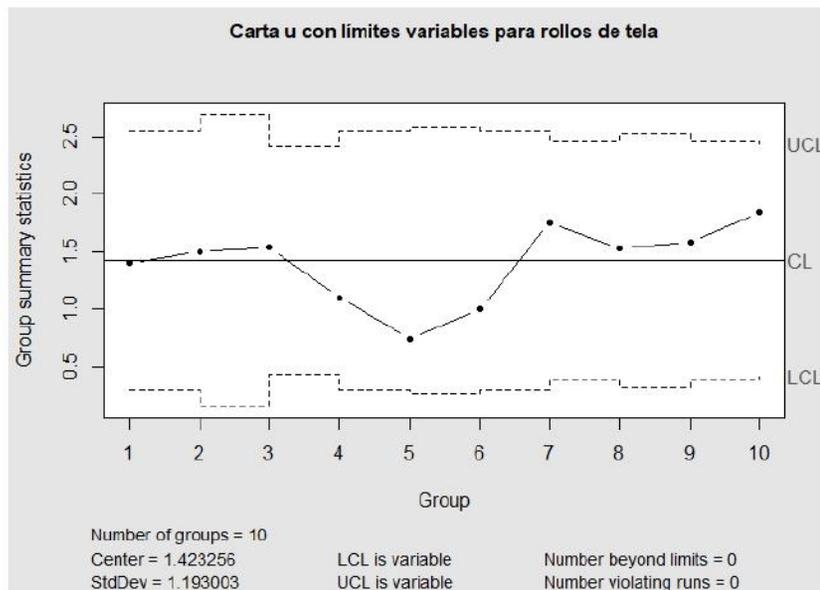


Figura 3: Carta u con límites variables

Puede notar en la *Figura 3* que para la elaboración de esta carta de control sólo se necesita cargar a *Rstudio*

los datos del número total de defectos o inconformidades y el tamaño de la muestra. Ya a partir de allí el software calculará los límites de control y de acuerdo al tamaño de la muestra asignado realizará las operaciones correspondientes que puede visualizar en la consola como se aprecia en la *Figura 4* si acude a los atributos de la función a partir del nombre asignado:

```
> rollos$statistics
      1      2      3      4      5      6      7      8
1.4000000 1.5000000 1.5384615 1.1000000 0.7368421 1.0000000 1.7500000 1.5238095
      9     10
1.5833333 1.8400000
```

Figura 4: Calculo de los puntos en la consola.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

A partir de la información dada, resuelva los siguientes incisos:

*En un hotel se ha venido llevando un registro de quejas de los clientes desde hace 15 semanas junto con el número de clientes por semana. Los datos se muestran en la Tabla 2:*

Tabla 3: Número de quejas para los clientes en 15 semanas.

Semana	Quejas	Cientes
1	11	114
2	15	153
3	5	115
4	14	174
5	16	157
6	11	219
7	10	149
8	9	147
9	10	131
10	10	91
11	10	112
12	11	158
13	30	244
14	11	111
15	11	120

1. Analice estos datos mediante una carta  $u$ .
2. Si este proceso estuviera en control estadístico, ¿eso quiere decir que la calidad es buena? Explique su respuesta.
3. Explique el significado de los límites de control obtenidos.
4. Si mejorara o empeorara la calidad, ¿cómo se daría cuenta a través de esta carta de control?

5. Tomando en cuenta únicamente el número de quejas, analízalas mediante una carta de control  $c$ .
6. ¿Por qué no se obtienen los mismos resultados con la carta  $c$  que con la  $u$ ?
7. ¿Cuáles serían los límites de control a usar en el futuro?

## 6. Referencias

- [1] D. V. Montgomery, Douglas C y Verbeek. *Control estadístico de la calidad*. Limusa Wiley, 2004.
- [2] Gutierrez Pullido, Humberto. *Calidad Total y Productividad Tercera Edición*. Mc Graw Hil, 2010

# PRACTICA 13: ANÁLISIS DE LA CAPACIDAD CP-CPK Y DESEMPEÑO DE UN PROCESO PP-PPK

## 1. Competencias a desarrollar

- ✓ Análisis de la capacidad y desempeño de un proceso gráficamente y mediante el cálculo estadístico a través del paquete *qcc*.
- ✓ Desarrollo de criterios para pruebas de normalidad en un conjunto de datos.
- ✓ Uso de gráficos de control estadístico en conjunto para establecer criterios en pro de mejoras del proceso.
- ✓ Desarrollo de pensamiento analítico en base a la información resultante.

## 2. Materiales requeridos

Elemento	Cantidad
Computador con sistema operativo Windows, Mac OS o Linux .....	1
Software de lenguaje de programación con enfoque en análisis estadístico <i>R</i> . .....	1
Interfaz de Ambiente de Desarrollo Integrado (IDE) <i>R Studio</i> .....	1

## 3. Teoría relacionada

### 3.1 Definición.

Existen casos en donde las cartas de control no son suficientes para poder realizar un análisis completo del control estadístico de la calidad. Si bien estas pueden establecer un panorama extenso del proceso, es requerido evaluar el comportamiento del producto final. Esto es importante debido a que puede un proceso estar bajo control estadístico y sin embargo no cumplir con los requerimientos demandados por el cliente.

Esta clase de problemáticas se abordan mediante el **análisis de la capacidad del proceso** lo cual mide la variabilidad de este de acuerdo a unas métricas establecidas. Estas métricas se evalúan a partir de un parámetro nominal  $N$  al que deben apuntar todos los productos elaborados y unos límites de desviación establecidos que se denominan **límites de especificación** y corresponden respectivamente al límite de especificación inferior y superior **LSL** y **USL** por sus siglas en inglés.

Todos estos conceptos se pueden apreciar en una distribución normal. En la *Figura 1* se evidencian las situaciones más comunes que se presentan al momento de analizar la capacidad de un proceso:

- ✓ El inciso **A)** muestra el caso ideal en el cual las unidades producidas se encuentran alineadas con el valor nominal que corresponde siempre a la media y además dentro de los límites de especificación.
- ✓ Por otra parte en el inciso **B)** se evidencia límites de especificación demasiado elevados. En menester en este caso ser más exigentes con el análisis y reducir el rango de estos límites.
- ✓ El ítem **C)** muestra el caso en el que si bien el proceso está centrado, este sobrepasa los límites de especificación establecidos. Esto es una señal de alarma ya que se están produciendo unidades no conformes a los requerimientos del cliente que acarrearán pérdidas considerables a la operación.

- ✓ Finalmente se aprecia el inciso **D)** en el cual la media de las unidades elaboradas no cumple con lo especificado, es decir, el proceso no está centrado.

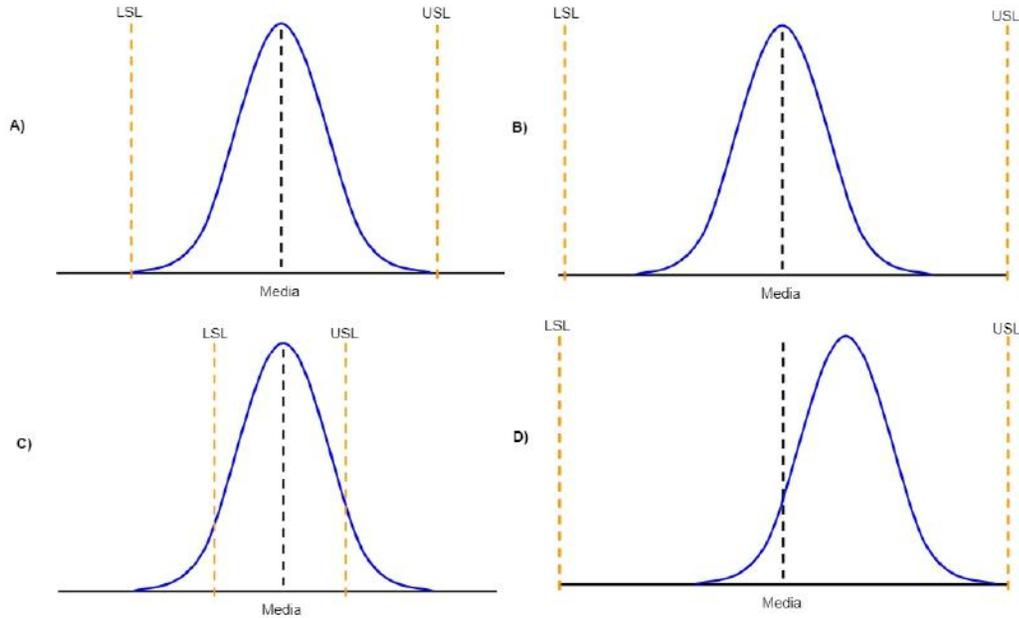


Figura 1: Gráficas de distribución normal con límites de especificación en el análisis de desempeño de procesos junto con los casos más comunes.

### Indices de capacidad de un proceso.

Una forma más precisa de determinar la capacidad del proceso es mediante un análisis cuantitativo mediante los **índices de capacidad** los cuales indican a través de un coeficiente qué tan bien está el performance del proceso.

El primer índice que se establece es el de **capacidad potencial**  $C_p$  el cual mide el rango o la extensión de las especificaciones en base a una dispersión seis sigma.[1]

$$C_p = \frac{USL - LSL}{6\sigma} \quad (1)$$

Donde  $\sigma$  corresponde al cociente entre el rango promedio  $\bar{R}$  de una muestra de datos y  $d_2$  una constante correspondiente al tamaño de la muestra:

$$\sigma = \frac{\bar{R}}{d_2} \quad (2)$$

Si bien es importante medir el proceso mediante el índice  $C_p$  debido a que indica si el proceso se encuentra dentro de las especificaciones, este tiene el inconveniente de dejar de lado el valor medio  $\mu$  lo cual no da la certeza del grado de centralidad del proceso. Para lograr este fin se calcula el **índice de la capacidad real del proceso**  $C_{pk}$  el cual si tiene este parámetro en cuenta. Para calcularlo se deben acudir a los **índices de capacidad para la especificación inferior y superior** respectivamente  $C_{pu}$   $C_{pl}$  los cuales calculan la distancia entre la media del proceso con respecto a algún extremo de las especificaciones. Como se están

calculando para los extremos inferior y superior respectivamente, sólo se utiliza una variación de 3 sigma. Estos cálculos se evidencian en las ecuaciones (3) y (4).

$$C_{pu} = \frac{USL - \mu}{3\sigma} \quad (3)$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma} \quad (4)$$

Una vez hallados estos dos parámetros, el índice  $C_{pk}$  se establece entre el mínimo de los indicadores anteriores.

$$C_{pk} = \text{Min} [C_{pu}, C_{pl}] \quad (5)$$

Este índice refleja un panorama más preciso del proceso, con lo que se deben trabajar de forma simbiótica para un efectivo análisis.

En la *Tabla 1* se especifica la clasificación que tiene cada valor de  $C_p$  con su respectivo dictamen. Estos datos también se utilizan para el índice  $C_{pk}$ :

*Tabla 1: Valores del índice  $C_p$  -  $C_{pk}$  y su indicación.*

Valor índice $C_p$ - $C_{pk}$	Clase o categoría del proceso	Dictamen (Si el proceso está centrado)
Mayor o igual a 2	Clase mundial	Calidad Seis sigma
Mayor a 1.33	1	Adecuado
Entre 1 y menor o igual a 1.33	2	Parcialmente adecuado
Entre 0.67 y menor o igual a 1	3	No adecuado para el trabajo. Requiere serias modificaciones.
Menor a 0.67	4	No adecuado para el trabajo. Requiere con necesidad imperiosa modificaciones

La antesala fundamental para el análisis de la capacidad de los procesos se debe establecer bajo los siguientes criterios:

1. Los datos tienen una distribución normal.
2. El proceso está bajo control estadístico.

Con lo cual, lo primero es establecer la normalidad de los datos ya que si se calculan los índices de capacidad en base a esto, se pueden cometer errores de magnitudes graves en lo que respecta a la incorrecta estimación de la variabilidad.

En lo que respecta al segundo inciso, es aquí donde cobra un papel importante el trabajo de la mano que se realice entre las cartas de control (las de variables son las más utilizadas, sin embargo las cartas de control por atributos también pueden ser un complemento adicional). En el caso de que el proceso no se encuentre bajo control estadístico, se deben calcular los **Índices de desempeño del proceso  $Pp$  y  $Ppk$**

## Indices de desempeño de un proceso

El *Grupo de Acción de la Industrial Automotriz* (AIAG, por sus siglas en inglés) fue formado en el año de 1991. Este grupo está integrado por los representantes de Ford, General Motors, Daimler/Chrysler y de la American Society for Quality Control. Uno de sus objetivos era estandarizar los requerimientos para los reportes de los proveedores y, en general de la industria. EL AIAG recomienda usar los índices de la capacidad del proceso  $C_p$  y  $C_{pk}$  cuando el proceso está bajo control estadístico, con la desviación estándar del proceso estimado  $\sigma$ . Cuando el proceso no está bajo control, el AIAG recomienda usar los **índices de desempeño**  $P_p$  y  $P_{pk}$  donde:

$$P_p = \frac{USL - LSL}{6S} \quad (6)$$

y  $S$  es la desviación muestral del total de los datos. [1]

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (7)$$

De tal manera, las ecuaciones (2), (3) y (4) se transforman en sus análogas (8), (9) y (10) que se muestran a continuación:

$$P_{pu} = \frac{USL - \mu}{3S} \quad (8)$$

$$P_{pl} = \frac{\mu - LSL}{3S} \quad (9)$$

De la misma manera, el índice  $P_{pk}$  se establece entre el mínimo de los indicadores anteriores.

$$P_{pk} = \text{Min}[P_{pu}, P_{pl}] \quad (10)$$

Muchos autores están en contra del uso de los índices de desempeño debido a que no es recomendable canalizar esfuerzos en optimizar un proceso que está alejado de su performance óptimo, sin embargo, resultan convenientes para poder visualizar los resultados respectivos en términos de corto y largo plazo. Esto toma relevancia debido a que por el lado de la ecuación (2) se calculan estos índices a partir de la desviación de una muestra mientras en la ecuación (7) se calcula toda la desviación de los datos.

## 4. Montaje y procedimiento

**Docente:** Indique al estudiante que ejecuten el software *Rstudio* y que realicen dos archivos mediante Excel lo cuales deberán importar a *Rstudio* cuyos datos corresponden a los expuestos en la *Tabla 2* de acuerdo a las siguientes características:

1. Un archivo en el que todos los datos estén en una sola columna llamada *Total*. El archivo tendrá "Normal" como nombre. Este archivo tiene el propósito de servir de base para la consecuente prueba de normalidad de los datos.
2. Un archivo con la misma estructura establecida en la *Tabla 2*. Estos datos servirán como insumo para la generación de las cartas de control requeridas y para el análisis de la capacidad del proceso. Nombre del archivo como "Datos".

! Recuerde que puede hacer la importación en la ventana superior derecha del software, dando click en el botón “Import Dataset” y luego dando click en “From Excel...”

**Estudiante:** Realice los dos archivos en Excel mencionados en el inciso anterior que *a posteriori* deberá importar a *Rstudio* en base a los datos del siguiente caso:

### Planteamiento del problema.

En la Tabla 2 se presentan los datos de la resistencia al estallamiento de 100 botellas de vidrio para refrescos de 1 litro.

Tabla 2: Datos de la resistencia al estallamiento de botellas

Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
265	205	263	307	220
268	260	234	299	215
197	286	274	243	231
267	281	265	214	318
346	317	242	258	276
300	208	187	264	271
280	242	260	321	228
250	299	258	267	293
265	254	281	294	223
260	308	235	283	277
200	235	246	328	296
276	264	269	235	290
221	176	248	263	231
334	280	265	272	283
265	262	271	245	301
280	274	253	287	258
261	248	260	274	337
250	278	254	274	275
278	250	265	270	298
257	210	280	269	251

! El en el ejercicio establecido en esta práctica, el cual corresponde al *Ejemplo 7-1* del capítulo 7 perteneciente al libro *Control estadístico de la calidad* de *Douglas Montgomery* solamente se indicó el valor del límite de especificación inferior *LSL* con un valor de 200 psi y se evaluó como un análisis unilateral. Sin embargo, como datos complementarios al ejercicio, han sido añadidos el límite de especificación superior *USL* de 300 psi y el valor nominal *N* de 267 psi.

## 4.2 Prueba de normalidad

Para realizar la prueba de normalidad, proceda con la importación a *Rstudio* del primer archivo nombrado como “Normal” Esto se hace de esta manera debido a que si bien existen comandos para agrupar columnas de un dataframe, lo que se requiere es colocarlas cada una debajo de la otra para poder verificar si los datos satisfacen una distribución normal. Una vez ha realizado esto, deberá escoger el tipo de prueba de acuerdo a

lo siguiente:

- ✓ **Shapiro-Wilk:** Esta prueba es utilizada para una cantidad de muestras menores a 50. Para esto deberá utilizar el comando `shapiro.test()`
  
- ✓ **Kolmogorov-Smirnov:** Funciona eficientemente para una cantidad de muestras mayores a 50. Esta es la que se utilizará en este caso. Para ello deberá hacer uso del paquete `nortest` y luego utilizar el comando `lillie.test()`. (Recuerde que si este paquete no se encuentra instalado en el terminal deberá hacerlo mediante el comando `install.packages` o bien directamente desde el área de entornos en la pestaña “Packages”). A continuación se muestra el código a ejecutar y su paso a paso:

### Elaboración del código para prueba de normalidad.

! No copie y pegue los textos que están entre comillas “ ” de esta guía directamente en área de edición de scripts o en la consola de *Rstudio*. Es recomendable escribir el texto adecuadamente o de lo contrario el código no funcionará correctamente. Si va a copiar y pegar, trate de hacerlo sólo con el texto y las comillas escríbalas en *Rstudio*.

```
library('qcc')
library('nortest')
attach(Normal)
lillie.test(Total)
hist(Total)
min(Total)
max(Total)
```

1. Realice el llamado de los paquetes `qcc` y `nortest`.
2. Haga uso de la función `attach` para poder llamar directamente la columna con toda la data.
3. Para verificar la normalidad de los datos realice la prueba de *Kolmogorov-Smirnov* debido a que el dataframe utilizado cuenta con 100 muestras. Al ejecutar el comando apreciará el resultado en la consola.

! El nombre de la columna utilizada en el archivo “Normal ” tiene el nombre de “**Total**”, es por esta razón que se coloca este en las líneas 4, 5, 6 y 7.

En *Rstudio* toda prueba de normalidad en realidad es una prueba de hipótesis la cual establece que si el valor del p-value es mayor a un  $\alpha$  de 0.05, el conjunto de datos presenta una distribución normal (lo cual corresponde a un nivel de confianza por defecto del 95%). En este caso el p-value obtenido de la prueba de normalidad del comando `lillie.test()` arroja un resultado de 0.07221 el cual es mayor al  $\alpha$  de 0.05 con lo cual, se acepta la hipótesis nula  $H_0$  de que el conjunto de datos tiene una distribución normal. En la *Figura 2* se puede apreciar el resultado de dicho test:

```

Exp<LSL 2.7%      Obs<LSL 3%
Exp>USL 14%      Obs>USL 10%
> lillie.test(Total)

      Lilliefors (Kolmogorov-Smirnov) normality test

data: Total
D = 0.084941, p-value = 0.07221

```

Figura 2: Prueba de normalidad Kolmogorov-Smirnov

- La distribución de los datos también se puede verificar de forma gráfica mediante el uso de un histograma. Esto no iguala la veracidad de una prueba de normalidad pero si establece un panorama sobre este. Elabore un histograma para la la columna “Total” con el comando `hist()`. También puede verificar los valores máximos y mínimos del conjunto de datos los cuales corresponden a 176 y 346 haciendo uso de los comandos `min()` y `max()` respectivamente.

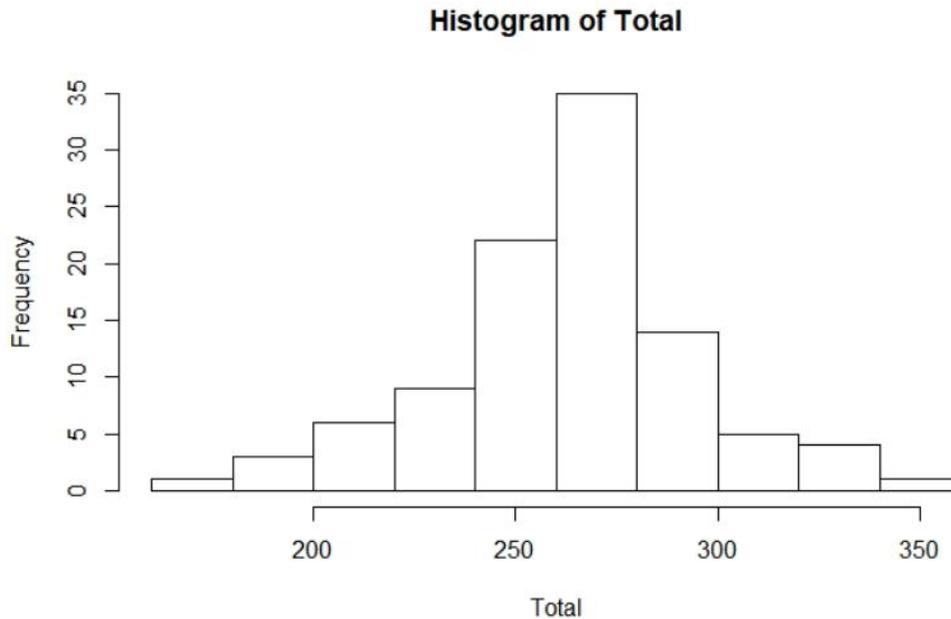


Figura 3: Histograma de los datos del archivo “Normal”.

### 4.3 Análisis de control estadístico

#### Elaboración de código para cartas de control.

Una vez se ha verificado la normalidad del proceso, se evaluará si este se encuentra en control estadístico. Para esto deberá elaborar las cartas  $\bar{X}$  y  $R$  para el segundo archivo “Datos”. Esto requiere asignar a cada código un nombre con el fin de establecer un objeto de clase `qcc` que se utilizará para el código de la capacidad del proceso.

En este caso se asignarán las letras *a* y *b* como nombres respectivamente a los códigos que generan las cartas de control  $\bar{X}$  y  $R$ . Los comandos a ejecutar tendrán la siguiente estructura:

```
a<-qcc(Datos,type = ‘‘xbar’’ ,std.dev = ‘‘UWAVE-R’’ ,confidence.level = 0.950,nsigmas = 3)
b<-qcc(Datos,type = ‘‘R’’)
```

Antes de elaborar los gráficos, recuerde que la función `qcc` tiene como forma general la siguiente estructura:

```
qcc(data, type, sizes, center, std.dev, limits, data.name, labels, newdata, newsizes,
newdata.name, newlabels, nsigmas = 3, confidence.level, rules = shewhart.rules,
plot = TRUE
```

### Otros argumentos para cartas de control.

La carta  $\bar{X}$  a parte de contar con los argumentos fundamentales **data** y **type** tendrá estos ajustes a los siguientes argumentos:

- ✓ **std.dev:** Un valor o un método el cual dictamina la desviación estándar de la carta de control. Recuerde Los métodos que dispone *Rstudio* para calcular la desviación estándar son:

*Tabla 3: Métodos disponibles de desviaciones estandar.*

Método	“xbar”	“R”	“S”
“UWARE-R”	Por defecto	Por defecto	No disponible
“UWARE-SD”	Disponible	No disponible	Por defecto
“MVLUE-R”			No disponible
“MVLUE-S”		No disponible	Disponible
“RMSDF”		No disponible	

Con lo cual, el tipo de desviación “UWAVE-R” se establecerá en este argumento.

- ✓ **confidence.level:** Es un valor que oscila entre 0 y 1 el cual indica el nivel de confianza del cálculo de los límites de control. En este caso se dejará por defecto.
- ✓ **nsigmas:** Número de sigmas especificado para los límites de control. Por defecto será un valor de 3, sin embargo se especificará para ambas cartas de control **nsigmas=3**. Este argumento acepta cualquier valor mayor a cero, sin embargo, lo recomendable es utilizar un valor de sigma de tres.

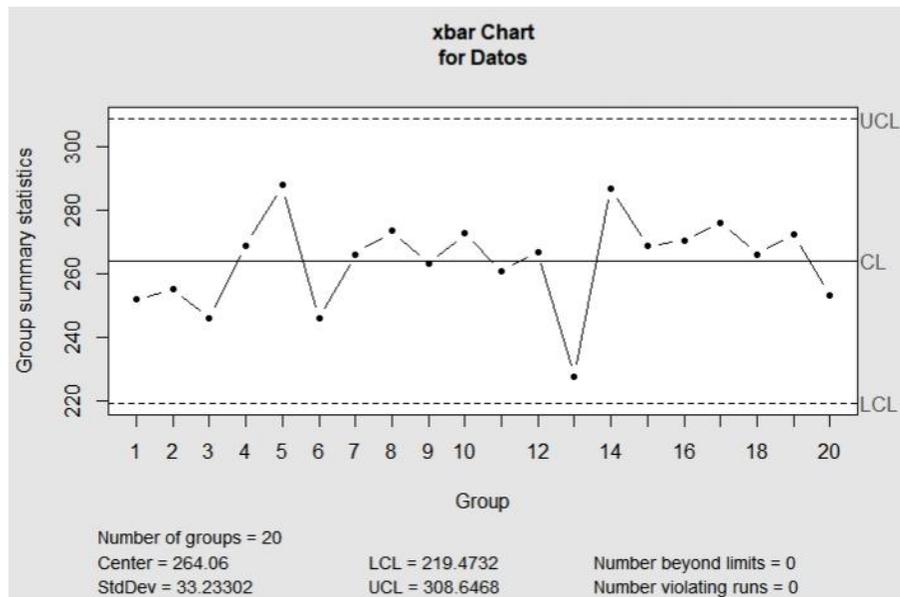


Figura 4: Carta X de medias para el conjunto de datos

En esta carta se evidencia un proceso dentro de los límites de control a pesar de que el punto 13 tuvo una caída significativa, acercándose mucho a LCL.

Por otro lado, la carta  $R$  solamente estará conformada por los argumentos esenciales de **data** y **type**. Si bien al momento de plotear el gráfico el proceso se encuentra bajo control estadístico, se evidencia una corrida en el punto 7, lo cual indica que algo puede que no esté resultando del todo bien para que los 7 primeros puntos presenten este comportamiento más aún cerca de la línea central. Tampoco se evidencia una estacionalidad en los puntos.

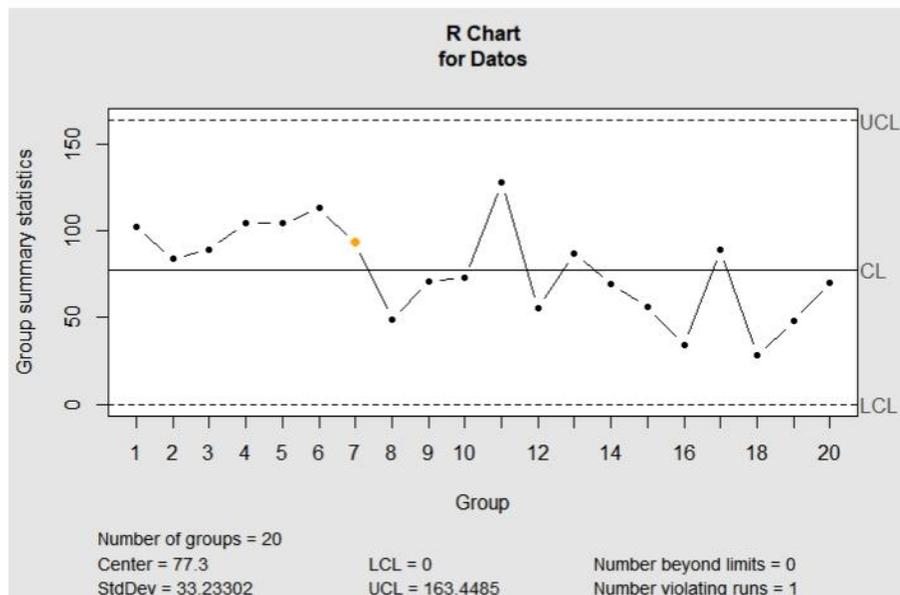


Figura 5: Carta R de rangos para el conjunto de datos

## 4.4 Análisis de capacidad del proceso

Elaboración del código para la función `process.capability`.

Con esto en mente, la función subsecuente `process.capability` es la encargada de realizar el análisis de la capacidad del proceso. La forma general de esta función corresponde a:

```
process.capability(object,spec.limits,  
target, std.dev, nsigmas, confidence.level, breaks ,  
add.stats = TRUE, print = TRUE, digits, restore.par = TRUE)
```

Esta comparte con la carta de control los argumentos **std.dev**, **nsigmas**, **confidence.level** (estos tres argumentos mencionados tendrán los mismos valores declarados en la carta  $\bar{X}$  ya que si no se colocan, por defecto asignará el valor de estos de la carta  $\bar{X}$ ); **add.stats** que permite visualizar las estadísticas en la parte inferior del gráfico; **digits** que controla el número de cifras significativas y **restore.par** que restaura los parámetros por defecto de la función `par` (si es que ha sido utilizada). Además de esto, debido a que se hace uso de un histograma utiliza de manera análoga a esta el argumento **breaks** el cual cambia el número de las barras a visualizar.

### Argumentos de la función `process.capability`

Los argumentos básicos para la función `process.capability` son:

- ✓ **object**: Un objeto de clase *qcc*. Recuerde que en la página 7 se ha asignado a los códigos de las cartas los nombres de *a* y *b*. Eso se hizo para poder utilizar esos nombres en este parámetro. Para este caso utilice el objeto de nombre *a*.
- ✓ **spec.limits**: vector binario que establece los límites de especificación. Para este caso dictamine 200 y 300 respectivamente.
- ✓ **target**: El valor nominal *N* al que se debe basar el proceso. Dictamine un valor de 267 psi.

De esta manera el código correspondiente deberá tener la siguiente estructura:

```
process.capability(a, spec.limits = c(200,300),  
target=267,breaks=20,nsigmas = 3,  
add.stats = TRUE,print = TRUE,restore.par = TRUE)
```

Con lo cual, el resultado se evidenciará en la *Figura 6*. Los índices de la capacidad del proceso se muestran en la parte inferior del gráfico  $C_p$  y  $C_{pk}$  presentan resultados muy pobres de acuerdo a lo observado en la *Tabla 1*. La administración deberá tomar muy seriamente cartas en el asunto ya que el producto si bien la media está relativamente centrada, se presenta una variabilidad evidentemente grande.

Los índices  $C_{pi}$ ,  $C_{pu}$ , se complementan con los porcentaje de esperados para cada límite de especificación  $Exp<LSL$  y  $Exp>USL$  que corresponden a 2.7% y 14% respectivamente los cuales son cifras garrafales ya que estarían esperando cantidades de partes por millón ppm de 27000 y 140000 unidades defectuosas.

Por otra parte se evidencian en sus magnitudes análogas los porcentajes observados de  $Obs<LSL$  y  $Obs>USL$  de 3% y 10%. lo cual corresponde a 30000 y 100000 partes por millón defectuosas. Preocupa el hecho de que existan demasiadas unidades por encima de límite superior de especificación ya que se estarían gastando

recursos en realizar botellas que no requieren tanta resistencia al embotellamiento. De manera similar, también se evidencia que se calcula el conocido *índice de Taguchi*  $C_{pm}$  sin embargo este también muestra un nivel bastante deplorable con lo cual, se deben extremar las medidas en pro de elevar la capacidad del proceso a niveles aceptables.

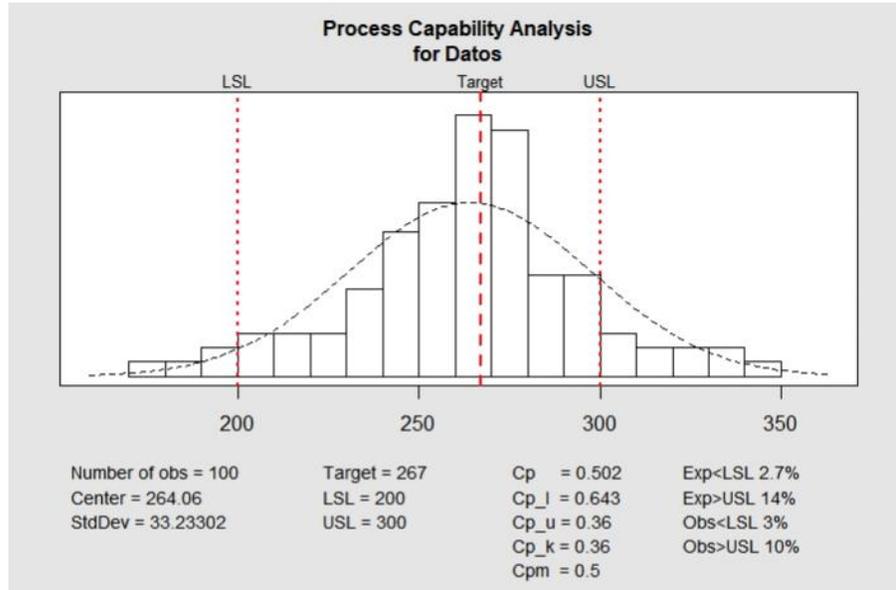


Figura 6: Gráfico resultante para el análisis de la capacidad del proceso.

Ya para finalizar, en la *Figura 7* se muestran en la consola las estadísticas obtenidas por la función `process.capability`. Se puede decidir si estas no son generadas en la consola a partir del argumento lógico `print`.

```
> process.capability(a, spec.limits = c(200,300))

Process Capability Analysis

Call:
process.capability(object = a, spec.limits = c(200, 300))

Number of obs = 100          Target = 250
      Center = 264.1         LSL = 200
      StdDev = 33.23         USL = 300

Capability indices:

      Value    2.5%   97.5%
Cp      0.5015  0.4317  0.5712
Cp_l    0.6425  0.5495  0.7355
Cp_u    0.3605  0.2913  0.4296
Cp_k    0.3605  0.2781  0.4429
Cpm     0.4619  0.3932  0.5304

Exp<LSL 2.7%      Obs<LSL 3%
Exp>USL 14%      Obs>USL 10%
```

Figura 7: Datos en la consola de los índices de capacidad.

## 4.5 Análisis de desempeño del proceso

### Cálculo de desviación.

Para calcular el desempeño del proceso se debe cambiar la desviación estándar asignando a un nuevo objeto llamado "Sigma\_Total" haciendo uso de la función `sd` en la cual se creará un vector y dentro de este hará uso de la función `as.matrix`. De esta manera el código tiene la siguiente estructura:

```
Sigma_Total <- sd(c(as.matrix(Datos)))  
Sigma_Total
```

! El el código anterior, la letra `c` hace referencia a un vector que tendrá el contenido descrito y no a un objeto en concreto.

Una vez realizado esto se procede a cambiar en el argumento `sd.dev` y se coloca el nombre del objeto. De ahí se realiza todo igual. De tal forma que el código tendría esta estructura:

```
process.capability(a, spec.limits = c(200,300), target=267,breaks=20,nsigmas = 3,  
add.stats = TRUE,print = TRUE,restore.par = TRUE,std.dev = Sigma_Total)
```

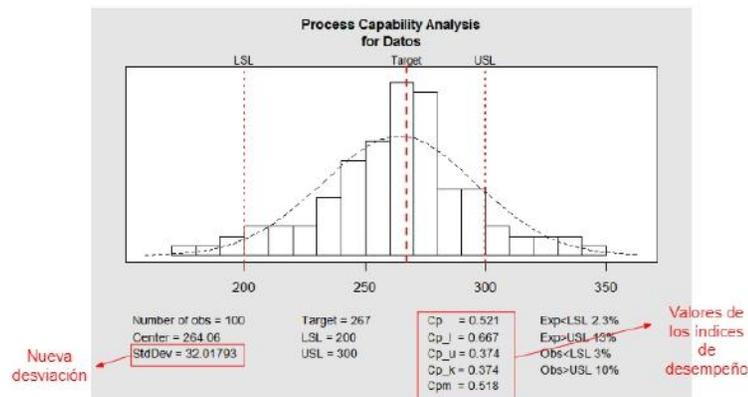


Figura 8: Análisis de los índices de desempeño.

```
Process Capability Analysis  
Cell:  
process.capability(object = b, spec.limits = c(200, 300), target = 267,  
s = 3, breaks = 20, add.stats = TRUE, print = TRUE, restore.par = TRUE)  
  
Number of obs = 100      Target = 267  
Center = 264.1          LSL = 200  
StdDev = 32.02         USL = 300  
  
Capability indices:  
Value      2.5%   97.5%  
Cp      0.5205  0.4481  0.5929  
Cp_l    0.6669  0.5716  0.7622  
Cp_u    0.3742  0.3040  0.4443  
Cp_k    0.3742  0.2906  0.4577  
Cpm     0.5184  0.4463  0.5903  
  
Exp<LSL 2.3%   Obs<LSL 3%  
Exp>USL 13%   Obs>USL 10%
```

Valores de los índices de desempeño con cuatro cifras significativas

Figura 9: Datos en la consola de los índices de desempeño.

En la *Figura 8* se evidencia el nuevo análisis de desempeño. Las etiquetas ( $Cp$ ,  $Cpi$ ,  $Cpu$ ,  $Cpk$ ) en la parte inferior del gráfico para capacidad no han cambiado, sin embargo las cifras, el factor fundamental, si. Esto se debe a que aún no se ha desarrollado en el paquete *qcc* un método directo como un comando de tipo `process.performance` para medir el desempeño con lo cual hay que acudir a este procedimiento. Por otra parte en la *Figura 9* evidencian respectivamente los valores en la consola con cuatro cifras significativas.

## 4.6 Análisis de capacidad y desempeño con el paquete *SixSigma*.

Existe un paquete llamado *SixSigma* el cual permite la elaboración de este procedimiento con unas pequeñas líneas de código y brinda un panorama más completo sin embargo, este presenta aún fallas, especialmente con los índices de desempeño lo hace que el análisis con este sea ineficiente.

### Elaboración del código con el paquete *SixSigma*.

Cada una de las líneas que se describirán a continuación realizan los cálculos en la consola. A continuación se explican cada una paso a paso y sus argumentos:

```
ss.ca.cp(Total, LSL = 200, USL = 300)
ss.ca.cp(Total, LSL = 200, USL = 300, ci = TRUE)
ss.ca.cpk(Total, LSL = 200, USL = 300)
ss.ca.cpk(Total, LSL = 200, USL = 300, ci=TRUE)
```

1. La primera línea calcula el índice  $Cp$ , mientras que la segunda calcula los intervalos de confianza en el que este índice oscila solamente dictaminando en el argumento `ci` como `TRUE`.
2. La penúltima línea calcula el índice  $Cpk$  y de la misma manera en el inciso anterior, añadiendo el argumento `ci` como verdadero se añaden los intervalos de confianza en la última línea.

Los argumentos de estas funciones corresponden a:

- ✓ **x**: Conjunto de datos para el proceso. En este caso será el nombre del archivo “Normal” dictaminado como “Total”.
- ✓ **LSL**: Límite de especificación inferior. Para este caso será de 200.
- ✓ **USL**: Límite de especificación superior. Igualmente para este argumento tendrá un valor de 300.
- ✓ **ci**: Argumento lógico el cual si se dictamina como verdadero lo que hará es calcular los intervalos de confianza del indicador.

Hasta este momento el paquete trabaja sin problemas sin embargo el inconveniente se presenta a la hora de ejecutar el argumento **LT**: Este argumento lógico si se dictamina como `TRUE` calculará el proceso a largo plazo, es decir, **calculará los índices de desempeño en donde se especifique**. (ss.ca.cp para índice  $Pp$  y ss.ca.cpk para índice  $Ppk$ .) El problema que se presenta es que lo que hace el paquete es prácticamente **copiar los valores de los índices de capacidad y pegarlos una vez se trata de ejecutar este código con este argumento**.

Este mismo inconveniente se presenta en la función donde se hace todo el análisis de forma gráfica `ss.study.ca`. Adicionalmente que el índice *Ppk* no lo muestra en el gráfico. Para poder hacer esta función se acude a la siguiente línea:

```
ss.study.ca(Total,LSL = 200, USL = 300, Target = 267, alpha = 0.05,
f.su = 'Análisis',f.main='Capacidad y Desempeño',
xLT=Total)
```

Esta comparte los mismos argumentos mencionados anteriormente y se adicionan:

- ✓ **Target:** Valor nominal *N*. Para este caso recuerde que es un valor de 276.
- ✓ **alpha:** Calcula el error tipo I para los intervalos de confianza. Este tendrá un valor de 0.05.
- ✓ **f.na.rm:** Si se dictamina como verdadero, los datos que sean NA serán eliminados.
- ✓ **f.main:** Argumento en el cual se le coloca título al gráfico. En este caso será `f.main='Capacidad y Desempeño'`.
- ✓ **f.sub:** Coloca un subtítulo en la parte inferior del gráfico. Para este caso se dictaminará como `f.sub='Análisis'`

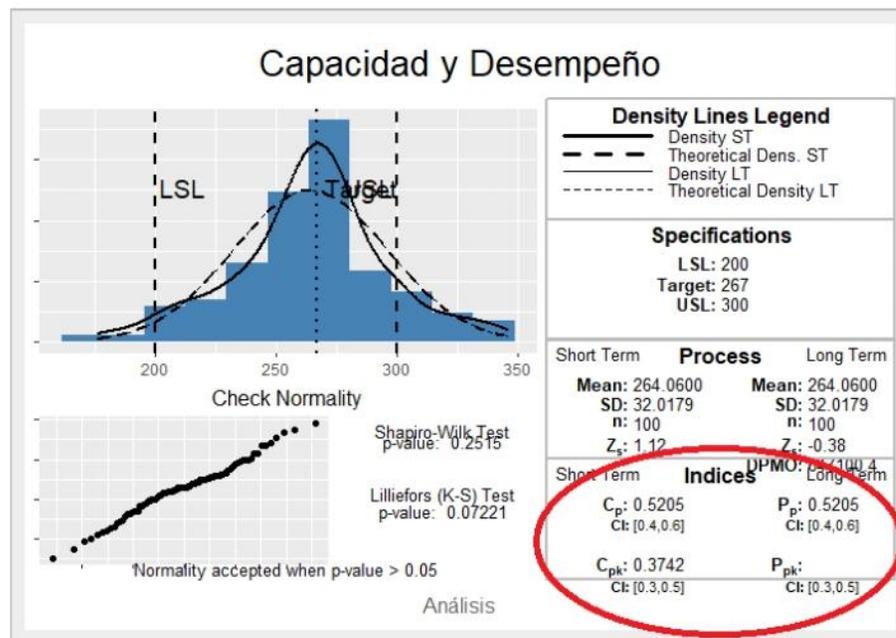


Figura 10: Gráfico resultante de análisis en el paquete *SixSigma*. Se evidencia en el círculo rojo las inconsistencias de este procedimiento.

De esta manera, en la *Figura 10* se evidencia el resultado. Este paquete realiza en unas cortas líneas un histograma, las pruebas de normalidad *Shapiro Wilk* y *Kolgomorov Smirnov*. Además añade un diagrama de dispersión y en la esquina inferior derecha se calculan los índices. En el círculo rojo se evidencia que los valores del índice *Cp* se copian en el índice *Pp* lo cual como vimos en el análisis realizado con el paquete *qcc* es erróneo. Por otra parte, el índice *Ppk* aparece en blanco.

De esta manera, el paquete *qcc* a pesar de tener que realizar varias líneas de código presenta mejores resultados. Es por esta razón que es la opción recomendada.

## 5. Evaluación

! La solución de la evaluación debe ser entregada en el informe de laboratorio. Tenga en cuenta que las respuestas deben estar relacionadas con lo observado durante la práctica.

A partir de la información resolver los puntos:

*Los siguientes datos representan las mediciones de viscosidad de los últimos tres meses de un producto lácteo. El objetivo es tener una viscosidad de  $80 \pm 10$  cps.*

84	81	77	80	80	82	78	83
81	78	83	84	85	84	82	84
82	80	83	84	82	78	83	81
86	85	79	86	83	82	84	82
83	82	84	86	81	82	81	82
87	84	83	82	81	84	84	81
78	83	83	80	86	83	82	86
87	81	78	81	82	84	83	79
80	82	86	82	80	83	82	76
79	81	82	84	85	87	88	90

1. Construya una gráfica de capacidad de este proceso (histograma con tolerancias) y genere una primera opinión sobre la capacidad.
2. Calcule la media y la desviación estándar y, considerando estos parámetros como poblacionales, estime los índices  $C_p$ ,  $C_{pk}$ ,  $C_{pm}$  y  $K$ , e intérpretelos con detalle.
3. Estime el porcentaje fuera de especificaciones.
4. Las estimaciones hechas en los dos incisos anteriores y las correspondientes interpretaciones se deben ver con ciertas reservas dado que son estimaciones basadas en una muestra. ¿Por qué se deben ver con reservas?

## 6. Referencias

[1] D. V. Montgomery, Douglas C y Verbeek. *Control estadístico de la calidad*. Limusa Wiley, 2004.

## ANEXO B. VIDEOS DE PRUEBAS A LAS GUÍAS

Una vez se terminó la etapa de elaboración de la guías, estas fueron sometidas a pruebas con varios participantes. Los videos en orden corresponden a:

1. Diagrama de tallo y Hoja e Histograma
2. Diagrama de caja
3. Diagrama de Pareto
4. Diagrama *Ishikawa*
5. Cartas *X-R*
6. Cartas *X-S* y carta *P*
7. Carta *X-One* y carta *np*
8. Carta *C*
9. Carta *U*
10. Análisis de capacidad y desempeño de procesos.

## BIBLIOGRAFÍA

- Alvarez, G. (Abril de 2019). *Significado*. Obtenido de Significado: <https://significado.com/comando-informatica/>
- Arce, K. A., & Morales, L. (2016). *Guías de laboratorio y programación del sistema CIM MPS500 para su uso en prácticas de métodos y tiempos en la Fundación Universitaria Los Libertadores*. Bogotá DC: Fundación Universitaria Los Libertadores.
- Barros, F. (s.f.). *IQCC*. Obtenido de IQCC: <https://flaviobarros.github.io/IQCC/reference/cchart.T2.2.html>
- Cano, E., Moguerza, J., Prieto, M., & Redchuk, A. (2018). *Six Sigma Tools for Quality Control And Improvement*. s.l: CRAN.
- Correa, J., & González, N. (2002). *Gráficos Estadísticos Con R*. Medellín: Universidad Nacional Sede Medellín.
- Flores, S. M. (2013). Desarrollo de una aplicación para graficos de control de procesos industriales. [http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_417.pdf](http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_417.pdf).
- Flores, M., Fernandez, R., Naya, S., Saavedra, T., & Javier. (2016). *Quality Control Review*. s.l: CRAN.
- Gutierrez, H. (2010). Calidad total y Productividad Tercera edición. En H. Gutierrez, *Calidad total y Productividad Tercera edición* (págs. 19-20). Ciudad de México: McGrawHill.
- Gutierrez, H. (2010). *Calidad Total y Productividad Tercera edición*. Ciudad de México: McGrawHill.
- Herrera, R., Hernandez, K., Figueroa, E., Ossa, D. I., & José. (2018). Aplicación del control estadístico multivariado para medir la capacidad del proceso de fabricación de resortes de compresión en acero inoxidable. *Scielo*, 58.
- Jensen, C., & Anderson, L. (1992). *Harvard Graphics 3: The Complete Reference*. McGrawHill.
- Lamport, L. (1986). *Latex: A Document Preparation System*. Addison-Wesley Publishing Company.

- Leek, J. (21 de Abril de 2016). *Write papers like a modern scientist (use Overleaf or Google Docs + Paperpile)*. Obtenido de Simplystats: <https://simplystatistics.org/2016/04/21/writing/>
- Lozano, M., & Ramos, I. (2000). *Ingeniería de Software y Bases de datos. Tendencias Actuales*. Albacete: Ediciones de la Ciudad de Castilla-La Mancha.
- Marroquin Escobar, E. S. (2017). *Simulación de control estadístico de la calidad con lenguaje R*. Guatemala : Universidad de San Carlos de Guatemala .
- Minitab. (2019). *support.minitab.com*. Obtenido de support.minitab.com: <https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/control-charts/how-to/multivariate-charts/multivariate-ewma-chart/before-you-start/overview/>
- Minitab. (s.f.). *pinzhi.org*. Obtenido de pinzhi.org: [http://www.pinzhi.org/Minitab/help/Mtbsg/qctimewt/cusu\\_010\\_cusum\\_chart\\_2\\_one\\_sided.htm](http://www.pinzhi.org/Minitab/help/Mtbsg/qctimewt/cusu_010_cusum_chart_2_one_sided.htm)
- Montgomery, D. (2004). *Control Estadístico De La Calidad*. México D.F: Limusa.
- mrbaronmaths. (06 de Septiembre de 2005). *mrbaronmaths*. Obtenido de <http://www.mrbaronmaths.com/resources/keystage3/data/>
- Ocaña, R. (2017). *La necesidad de convertir la Estadística en profesión regulada*. Granada.
- Paradis, E. (3 de Marzo de 2003). *R para Principiantes*. Obtenido de [https://cran.r-project.org/doc/contrib/rdebut\\_es.pdf](https://cran.r-project.org/doc/contrib/rdebut_es.pdf)
- Pastrana, Y. (2017). *Guías de laboratorio*. Bogotá: Fundación Universitaria Los Libertadores.
- RStudio. (s.f de s.f de s.f). *Rstudio*. Obtenido de Rstudio: <https://rstudio.com/products/rstudio/>
- Saez, A. J. (2010). *Metodos Estadísticos con R y R Commander*. España: Universidad de Jaen .
- Santana, J. S., & Farfán, E. M. (2014). *El arte de programar en R*. Mexico: Unesco. Obtenido de [https://cran.r-project.org/doc/contrib/Santana\\_El\\_arte\\_de\\_programar\\_en\\_R.pdf](https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf)
- Santos, E. (2016). *Multivariable Statistical Quality Control*. CRAN.

- Scrucca, L., Snow, G., & Boomfield, P. (2017). *Quality Control Charts*. s.l.: CRAN.
- Sharelatex. (s.f de s.f de s.f). *Sharelatex*. Obtenido de Sharelatex:  
<https://cutt.ly/MdFgo8a>
- Team, T. R. (2019). *R: A Language and Enviroment for Statistical Computing Version 3.6.2*. s.l: R Foundation for Statistical Computing.
- Technopedia. (31 de Marzo de 2020). *Technopedia*. Obtenido de Technopedia:  
<https://www.techopedia.com/definition/4356/software>
- Triola, M. (2004). *Estadística*. Ciudad de Mexico: Pearson.
- Tukey, J. W. (1993). *Exploratory Data Analysis: Past, Present and Future*. Princeton: Princeton University.
- vBulletin. (s.f de s.f de s.f). *vBulletin*. Obtenido de vBulletin:  
[https://forum.vbulletin.com/help?faq=vb3\\_board\\_usage#faq\\_vb3\\_forums\\_threads\\_posts](https://forum.vbulletin.com/help?faq=vb3_board_usage#faq_vb3_forums_threads_posts)
- Venables , B., & Smith, D. M. (Febrero de 1990). *Introduccion a R*. Obtenido de  
<https://cran.r-project.org/doc/manuals/r-release/R-intro.html#Preface>
- Wickham, H. (2014). *Advanced R*. CRC Press.