



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

Análisis de control estadístico de calidad usando el software R.

TESIS

Que para obtener el título de
Ingeniero Industrial

P R E S E N T A

Mejía Cervantes Karen Evely

DIRECTORA DE TESIS

MI. Ann G. Wellens Purnal



Ciudad Universitaria, Cd. Mx., 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

A mis padres Eugenia Cervantes Sánchez y Moises Mejía Reyes que me han guiado por la vida con amor y paciencia, gracias al apoyo brindado he logrado terminar mis estudios profesionales por lo cual viviré eternamente agradecida, solo quiero hacer mención que cualquier logro mío, es un logro suyo, gracias por confiar siempre en mí, los amo con todo mi corazón.

A mis hermanos Moises Mejía Cervantes y Marlene Mejía Cervantes por ser mis compañeros de vida y superación incansable por su comprensión y confianza, les dedico este logro de todo corazón.

A la Universidad Nacional Autónoma de México, por haberme permitido ser alumna de esta gran institución y permitirme ahora, formar parte del grupo de profesionistas de este país.

A mi asesora Ann. Wellens Purnal por la paciencia y el tiempo dedicado durante la elaboración de este trabajo y transmitirme los conocimientos necesarios para poder cruzar la meta y poder concluir mis estudios profesionales.

INDICE

I.	Introducción.....	7
II.	Justificación del trabajo.....	7
III.	Objetivo	8
IV.	Objetivos específicos.....	8
1	Marco teórico	9
1.1	Antecedentes	9
1.2	R en la calidad y análisis estadístico	11
	Herramientas básicas para el mejoramiento de la calidad	14
1.3.1	Hoja de verificación.....	15
1.3.2	Tormenta de ideas	15
1.3.3	Un diagrama causa – efecto	16
1.3.4	Histograma	17
1.3.5	Diagrama de Pareto	17
1.3.6	Diagrama de dispersión.....	18
1.3.7	Gráfica de control	20
1.4	Muestreo de datos	20
1.5	Prueba de hipótesis	21
1.6	Estudio de capacidad del proceso	23
1.7	Diagramas de control.....	27
1.8	Muestro de aceptación	32
1.9	¿Qué es R?.....	33
1.9.1	Características.....	34
1.9.2	Descarga e instalación de R	35
1.9.3	Rstudio	35
2	Funcionamiento general de R y manejo de funciones	37
2.1	Uso de R	37
2.1.1	Ayuda	37
2.2	Tipos de variables y su asignación	38
2.2.1	Como manejar distribuciones de probabilidad	38
2.3	Leer datos desde un archivo	41
2.3.1	Graficar.....	42

2.3.2	Paquetes y complementos	44
3	Generación de los códigos aplicables.....	46
3.1.	Descripción del proceso hipotético analizado	46
3.1.1.	Generación de histogramas.....	50
3.1.2.	Prueba de hipótesis	51
3.2.	Herramientas básicas de la calidad	55
3.2.1.	Diagrama de Pareto	55
3.2.2.	Diagramas de dispersión	56
3.3.	Estudio de capacidad del proceso	58
3.3.1	Estimación de μ_x y σ_x	59
3.3.2	Representación gráfica de los resultados	62
3.4.	Diagramas de control.....	63
3.4.1.	Diagramas de control empleando la distribución binomial	67
3.4.2.	Diagramas de control empleando la distribución de Poisson	71
3.4.3.	Diagramas CUSUM y EWMA.....	74
4	Análisis de control estadístico e interpretación de códigos generados	80
4.1	Prueba de hipótesis	81
4.2	Índices de capacidad.....	82
4.3	Diagramas de control Xpr-R y Rango	83
4.4	Diagramas de control CUSUM y EWMA.....	85
4.5	Diagramas de control empleando la distribución binomial con n variable	86
4.6	Diagramas de control empleando la distribución de Poisson	87
5	Conclusiones	89
6	Bibliografía.....	91
7	Anexos	92
7.1	Minitab.....	92
7.2	Excel	95

Índice de figuras

Figura 1.1 Factores que influyen en la calidad	11
Figura 1.2 Esquema de un proceso	12
Figura 1.3 Esquema de diagrama causa-efecto	16
Figura 1.4 Tipos de distribución de frecuencias	17
Figura 1.5 Diagrama de Pareto de quejas	18
Figura 1.6 Patrones de correlación.	19
Figura 1.7 Diagrama de flujo para selección de gráficos de control	28
Figura 1.8 Pantalla inicial de Rstudio	36
Figura 2.1. Curva de distribución normal	40
Figura 2.2. Pantalla de R Studio para instalar Packages.....	44
Figura 2.3. Instalación de paquete en R.....	45
Figura 2.4. Espacio de trabajo con el paquete instalado	45
Figura 3.1 Histograma de peso de paquetes de pasta de la línea 1	51
Figura 3.2 Histograma de peso de paquetes de pasta de la línea 1 personalizado.....	51
Figura 3.3 Gráfico de la distribución normal	53
Figura 3.4 Gráfico con la zona de rechazo de la hipótesis nula	54
Figura 3.5 Gráfico con la zona de rechazo y aceptación de la hipótesis nula	54
Figura 3.6 Gráfico con la zona de rechazo y aceptación de la hipótesis nula y Z de prueba	55
Figura 3.7 Datos del diagrama de Pareto.....	56
Figura 3.8 Diagrama de Pareto	56
Figura 3.9 Diagrama de dispersión	58
Figura 3.10 Gráfico de distribución normal.....	59
Figura 3.11 Límites de especificación línea 1	63
Figura 3.12 Diagrama de control X-R	67
Figura 3.13 Diagrama de control fracción disconforme (p)	70
Figura 3.14 Diagrama de control estandarizado fracción disconforme	71
Figura 3.15 diagrama de control para disconformidades por unidad (u)	73
Figura 3.16 Diagrama de control estandarizado para las disconformidades por unidad	74
Figura 3.17 Diagrama de control CUSUM	77
Figura 3.18 Diagrama de control EWMA.....	79
Figura 4.1 Histograma generado con Rstudio con límites de especificación y con 11 clases. a) línea 1, b) línea 2, c) línea 3.....	80
Figura 4.2 Distribución normal de prueba de hipótesis. a) línea 2, b) línea 3.....	81
Figura 4.3 Diagramas de control a) \bar{X} y b) R para los datos de la línea 2	84
Figura 4.4 Diagramas de control a) \bar{X} y b) R para los datos de la línea 3	85
Figura 4.5 Diagramas de control a) CUSUM y b) EWMA para los datos de la línea 2.....	85
Figura 4.6 Diagramas de control a) CUSUM y b) EWMA para los datos de la línea 3.....	86
Figura 4.7 Diagrama de control a) fracción disconforme (p) b) <i>normalizada</i> para los datos de la línea 2.....	87
Figura 4.8 Diagrama de control a) fracción disconforme (p) b) <i>normalizada</i> para los datos de la línea 3.....	87

Figura 4.9 Diagrama de control a) disconformidades por unidad (<i>u</i>) b) <i>normalizada</i> para los datos de la línea 2	88
Figura 4.10 Diagrama de control a) disconformidades por unidad (<i>u</i>) b) <i>normalizada</i> para los datos de la línea 3	88
Figura 7.1 Estadísticos descriptivos	92
Figura 7.2 Histograma	92
Figura 7.3 prueba de hipótesis (Z de una muestra)	92
Figura 7.4 Informe de capacidad del proceso	92
Figura 7.5 Diagrama de control Xbarra-R	93
Figura 7.6 Diagrama de control de tiempo ponderado EWMA	93
Figura 7.7 Diagrama de control fracción disconforme (p)	93
Figura 7.8 Diagrama de control disconformidades por unidad (<i>u</i>)	94
Figura 7.12 Diagrama de control CUSUM y EWMA	96
Figura 7.13 Diagrama de control a) Fracción disconforme (p) b) Normalizada	96

Índice de tablas

Tabla 1.1 Comparación de características entre softwares	10
Tabla 1.2 Historia de la calidad	13
Tabla 1.3 Hoja de verificación	15
Tabla 1.4 Quejas registradas	18
Tabla 1.5 Reglas de decisión de una hipótesis estadística	22
Tabla 1.6 Clasificación de los procesos de acuerdo con su capacidad	24
Tabla 1.7 Fórmulas para el cálculo de límites de control para variables	29
Tabla 1.8 Fórmulas para el cálculo de límites de control para atributos	30
Tabla 2.1 Operadores en R	37
Tabla 2.2 Tipos de distribuciones en R	39
Tabla 2.3 Descripción de los prefijos de la distribución normal en R	39
Tabla 2.4 Funciones graficas en R	43
Tabla 3.1 Muestreo de la línea 1,2 y 3	46
Tabla 3.2 Defectos de calidad	55
Tabla 3.3 Número de defectos	57
Tabla 3.4 Factores para la construcción de los diagramas de control	65
Tabla 3.5 Muestreo de defectos y unidades defectuosas	68
Tabla 4.1 Resumen de estadísticas generadas por Rstudio	80
Tabla 4.2 Valores de prueba de hipótesis	81
Tabla 4.3 Resumen de índices de capacidad por Rstudio	82
Tabla 4.4 Relación Cp y Cpk	83
Tabla 4.5 Probabilidad de paquetes defectuoso	83

I. Introducción

En la actualidad, una empresa de manufactura o de servicios difícilmente puede subsistir sin realizar un esfuerzo deliberado por mejorar la calidad de sus productos o servicios. Para lo anterior, las empresas tienen a su alcance distintos paquetes disponibles para un correcto análisis de los procesos de calidad involucrados en su sistema, varios de ellos basados en la estadística, tales como excel, minitab, SPSS, entre otros. Estos paquetes han sido desarrollados dentro de empresas lucrativas por lo que tienen un costo asociado a su uso. Sin embargo, debido al desarrollo rápido de las tecnologías de información y el creciente interés de la comunidad académica de compartir los conocimientos generados en forma libre, varios grupos académicos han buscado desarrollar herramientas gratuitas de uso libre. Un ejemplo es el paquete R; la ventaja de usar R es que es un software libre y gratuito en el cual se pueden manipular los datos para encontrar el análisis más adecuado que refleje el estado actual de una empresa con base en números.

R es una opción más en el mercado y viable para el análisis de datos estadísticos en el control de la calidad, tanto en procesos de manufactura, así como en procesos transaccionales. El simple hecho de ser un software libre facilita su acceso a diferentes sectores industriales o de servicios, esto es porque no se necesitan licencias para poder utilizarlo por lo cual los costos incurridos por el uso de este son casi nulos en comparación con otras opciones disponibles en el mercado.

De acuerdo con lo mencionado anteriormente, en este trabajo se busca dar a conocer lo versátil, flexible y potente que es el software, específicamente en el tema de control de la calidad.

II. Justificación del trabajo

Actualmente, el análisis estadístico de control de procesos es fundamental para cualquier empresa; este se lleva a cabo a través de distintos paquetes que sirven para la manipulación y presentación de datos. R, aunque es muy potente, versátil y solicitado por las empresas internacionales y también mexicanas, así como el estándar de uso para análisis estadístico en universidades muy reconocidos a nivel internacional, sigue teniendo poco uso entre los estudiantes de la Facultad de Ingeniería de la más grande casa de estudios en América Latina, la UNAM. La Facultad, sus profesores y sus estudiantes se deben involucrar más en la utilización y exploración de softwares que están teniendo un enorme auge en distintos sectores del conocimiento como lo es R, para seguir a la vanguardia.

El software R se ha convertido en una herramienta muy eficiente y el uso adecuado del mismo sirve de gran ayuda y ahorra tiempo de proceso en los datos. El deber de todo ingeniero es innovar y buscar alternativas de solución, por lo cual en este trabajo se

pretende mostrar la aplicabilidad de R dentro de la ingeniería industrial y dar a conocer la gran potencia que tiene el mismo desde análisis de datos muy simples hasta análisis de datos muy robustos, al desarrollar un código de programación para un análisis de control estadístico de la calidad.

III. Objetivo

Ejemplificar el análisis estadístico del control de la calidad en un proceso industrial o de servicios por medio de la programación en R tomando en cuenta paquetes existentes, así como la elaboración de códigos programáticos propios como apoyo al aprendizaje de los estudiantes de la carrera de ingeniería industrial.

IV. Objetivos específicos

- Proporcionar conocimiento acerca del software R y el impacto que está teniendo su uso en la manipulación de datos.
- Utilizar herramientas en R para el análisis de una base de datos de un proceso
- Realizar un código de programación en R para presentar el análisis de los datos
- Identificar las herramientas que se utilizan en el análisis de control estadístico de calidad y la interpretación de estas.

1 Marco teórico

1.1 Antecedentes

Actualmente, el uso de R ha presentado una tendencia positiva debido a que es flexible y potente; está diseñado para operar de la manera en la cual se está pensando con base en un lenguaje de programación, dicho lenguaje se puede diseñar de acuerdo con las necesidades del usuario.

Aunque también se usa ampliamente para aplicaciones en otras ramas de las matemáticas, R está enfocado al análisis estadístico de datos, para lo que el usuario puede manipular sus datos de la manera que más le convenga. Es importante mencionar que la estadística es una ciencia aliada a la investigación científica. Neyman (1955) proclamó a la estadística como "la sirvienta para todas las ciencias", ejemplificando su importancia en el estudio de la existencia y evolución, medicina, psicología, industria y astronomía.

Con el paso del tiempo, el uso de la tecnología ha revolucionado a la estadística e inversamente la estadística ha revolucionado a la tecnología. La generación de datos está creciendo exponencialmente; esto se debe a la gran cantidad de información que se produce procedente de un sinfín de dispositivos que forman parte de la vida cotidiana, desde ingresar a una página en internet, hacer alguna compra por internet, pagar una tarjeta de crédito, etc. Las herramientas tradicionales para almacenar, organizar y analizar datos llegan a ser insuficientes, lo que genera un término muy popular hoy en día llamado Big Data, el cual no tiene una definición oficial; sin embargo, una empresa consultora y de investigación de las tecnologías de la información con sede en Stamford definió el término de la siguiente manera: "Big Data son recursos de información de alto volumen y alta velocidad y/o de gran variedad que requieren nuevas formas de procesamiento para permitir una mejor toma de decisiones, nuevos descubrimientos y optimización de procesos" (Gartner, 2011). La explosión de este fenómeno está transformando la manera en que se conduce una investigación, adquiriendo habilidades en el uso de Big Data para resolver problemas complejos relacionados con el descubrimiento científico, investigación ambiental y biomédica, educación, salud, seguridad nacional, entre otros.

Cabe mencionar que en la actualidad se dispone de paquetes de cómputo estadístico que se pueden usar para manejar un volumen de datos considerablemente grande, así como una gran cantidad de variables; sin embargo, estos paquetes tienen un costo asociado, así como también son limitados en algunas disciplinas, por lo cual R al ser un software libre y de código abierto facilita el acceso al mismo para ser empleado en áreas cuantitativas de diversas disciplinas.

En México, universidades e instituciones como el Instituto Tecnológico Autónomo de México (ITAM), Centro de Investigación y de Estudios Avanzados (CINVESTAV), Universidad de Quintana Roo, Centro de Investigación en Matemáticas (CIMAT), así como la Universidad Nacional Autónoma de México (UNAM), por mencionar algunas, se está usando este software como herramienta para el análisis de datos estadístico en distintas disciplinas. En la UNAM, en algunas facultades como la Facultad de Psicología, Facultad de Ciencias, Facultad de Economía y en la Facultad de Ingeniería, algunos investigadores y profesores emplean este software para realizar pruebas de hipótesis, ajustes de modelos y análisis de diseños experimentales complejos entre otros tipos de análisis por lo flexible y potente que resulta ser R. Sin embargo, en la Facultad de Ingeniería su uso todavía no es muy generalizado.

Tomando en cuenta que este trabajo está centrado en el control estadístico de calidad, durante el desarrollo de este se realizará una comparación de aspectos generales entre los programas estadísticos que más comúnmente son usados para realizar el análisis como son excel y minitab, así como el software que se está proponiendo usar, que es R.

En la tabla 1.1. se comparan los tres softwares en aspectos como manipulación de datos, costo, amigabilidad con el usuario, sistema operativo, entre otros.

Aspecto	Excel	Minitab	R
Interacción con el usuario	Buena	Buena	Baja-Regular
Manipulación de datos	Buena	Buena	Buena
Calidad de gráficos	Buena	Buena	Excelente
Control de procesos	Excelente	Excelente	Excelente
Costo	\$1,299 (anual)	\$1,495 (anual)	Gratis
Código fuente disponible	No	No	Sí
Variedad de análisis estadísticos	Buena	Buena	Excelente
Soporte técnico	Bajo	Bajo	Alto
Sistema operativo	Windows Macintosh Linux	Windows Macintosh Linux	Windows Macintosh Linux

Tabla 1.1 Comparación de características entre softwares. Elaboración propia

De acuerdo con la tabla 1.1, R comparte su código fuente con la comunidad que tiene interés en desarrollar códigos de programación, la cual aporta otras formas de resolver problemas. Cabe mencionar que cada vez más usuarios se están involucrando dentro de la comunidad que usa R, por lo que cada uno de esos usuarios recibe conocimiento que se encuentra disponible y al mismo tiempo aporta otro conocimiento.

1.2 R en la calidad y análisis estadístico

Aunque R se usa en un campo más amplio que solo la estadística, es en esta rama del conocimiento que ha tenido más auge, debido a que fue en esta rama en donde nace este lenguaje, cuya característica principal es que forma un entorno de análisis estadístico para la manipulación de datos, su cálculo y creación de gráficos.

La estadística está formada por un conjunto de técnicas y conceptos orientados a la recolección y análisis de datos tomando en cuenta la variación en los mismos. Por su parte, el control estadístico de la calidad es la aplicación de técnicas estadísticas al control de calidad.

El análisis de control estadístico de calidad es importante para las empresas y/u organizaciones ya que ellas existen para proveer de un producto material o inmaterial, un bien o un servicio a sus clientes. Para cumplir con las expectativas de los clientes se necesita tener un proceso capaz y consistente, lo cual por medio del análisis estadístico se puede medir y controlar.

Es importante definir el concepto de calidad: todo el mundo habla de la calidad de los productos, pero ¿Qué es calidad?, para responder a esta pregunta se buscaron varias definiciones, por ejemplo, Juran (1990) dice que: “Calidad es que un producto sea adecuado para su uso. Así, la calidad consiste en la ausencia de deficiencias en aquellas características que satisfacen al cliente”, mientras que de acuerdo con Gutiérrez (2009) “Calidad es la creación continua de valor para el cliente”. Para cumplir con la calidad de un servicio o producto es importante tener claro 4 factores, como se muestran en la figura 1.1 recuperada de Gutiérrez (2009); estos son competitividad, productividad, variabilidad y mejora, para que una empresa cumpla con los requerimientos del cliente.

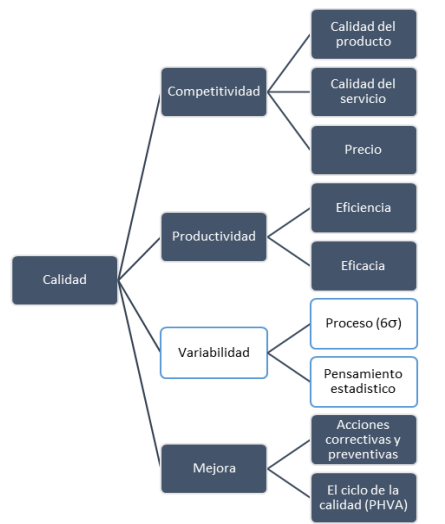


Figura 1.1 Factores que influyen en la calidad. Recuperado de Gutiérrez (2009)

De acuerdo con Gutierrez (2009), los productos son resultado de un proceso, el cual es un conjunto de actividades entrelazadas o interrelacionadas que reciben determinados insumos (entradas) que son transformados en un resultado (salidas) o en un producto. En la figura 1.2, recuperada de Gutierrez (2009), se muestra cómo se compone un proceso tomando en cuenta la voz del cliente, así como el diseño y la planeación para después pasar por varias etapas dentro del proceso para que se pueda obtener un producto o servicio.

Este trabajo aborda específicamente el aspecto de la variabilidad de un proceso industrial, así como su disminución y consiguiente mejora del proceso.

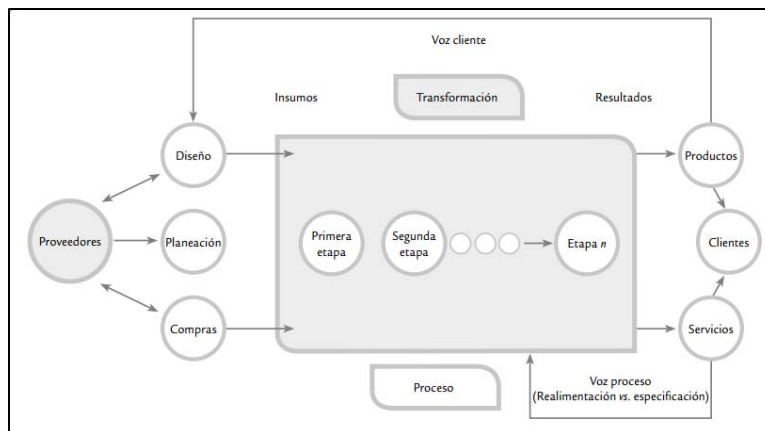


Figura 1.2 Esquema de un proceso. Recuperado de Gutierrez (2009)

Cuando se tiene mala calidad en las diferentes actividades hay equivocaciones y fallas de todo tipo, por ejemplo:

- Reprocesos, desperdicios y retrasos en la producción
- Reclamaciones
- Paros y fallas en el proceso
- Una inspección excesiva para tratar que los productos de mala calidad no salgan al mercado
- Re-inspección y eliminación de rechazo
- Más capacitación, instrucciones y presión a los trabajadores
- Clientes insatisfechos y pérdidas de ventas.

Con base en todas estas equivocaciones de la calidad, Juran (1990) menciona que: “Las necesidades humanas de calidad han existido desde el alba de la historia; sin embargo, los medios para satisfacer esas necesidades – los procesos de gestión para la calidad – han sufrido unos cambios amplios y continuos”.

En la tabla 1.2 se muestra cómo han ido evolucionando los procesos de gestión para la calidad con el paso del tiempo dentro de un mundo tan competitivo en donde se busca

incrementar cada día la calidad de los productos o servicios y una vez que el cliente consume y aprueba se busca mejorar la calidad para evitar perder clientes.

CONDICIÓN	ESTRATEGIAS ADOPTADAS EN LA GESTIÓN PARA LA CALIDAD
Obtención de alimentos	Inspección en la recepción por los consumidores
División del trabajo: proveedores de alimentos	Inspección por los consumidores en los mercados de los pueblos
Primeros fabricantes; nacimiento de los artesanos del pueblo	Confianza en la destreza y reputación de los artesanos
Expansión del comercio más allá de los límites del pueblo	Especificaciones: materiales, procesos, productos; controles de la exportación; auditorías
Los gremios	Especificaciones escritas; medidas, instrumentos, laboratorios de ensayo, extensión de la inspección; normalización
La Revolución Industrial	Departamentos centrales de inspección
El sistema Taylor	Departamentos de garantía de calidad; ingeniería de calidad, ingeniería de fiabilidad
Crecimiento del volumen y de la complejidad	Formación en control estadístico de la calidad.
Segunda guerra mundial	Organización y procesos especiales para proteger la sociedad; auditorías
La vida detrás de los diques de la calidad	Altos directivos personalmente implicados Formación en la gestión para la calidad extendida a todas las funciones La mejora de la calidad en un ritmo continuo revolucionario Círculos de CC
La revolución japonesa de la calidad (estrategias japonesas)	Esfuerzos por restringir las importaciones Muchas estrategias sometidas a prueba

Tabla 1.2 Historia de la calidad. Recuperado de Juran (1990)

Teniendo en cuenta los conceptos anteriores, cabe mencionar que un proceso no puede ser mejorado o controlado sin conocer su estado actual y sin haber medido antes sus características, para lo cual existen varias herramientas que se utilizan para el aseguramiento de la calidad; estas sirven para medir el desempeño del proceso industrial o del servicio. Estas herramientas serán detalladas en los siguientes subtemas.

1.3 Herramientas básicas para el mejoramiento de la calidad

Las herramientas más utilizadas por las organizaciones en su camino hacia la calidad son siete y fueron propuestas por Kaoru Ishikawa en 1968; estas herramientas tienen su fundamento en métodos estadísticos tradicionales que se enfocan a la solución de problemas comunes, como pueden ser los defectos en los productos. Estos defectos son creados a veces por la enorme variabilidad encontrada en los procesos, teniendo su raíz en la variación de los materiales, maquinaria, métodos de trabajo, mano de obra e inclusive el medio ambiente. La siguiente información se obtuvo de Gutierrez (2009), (Montgomery, 2016).

Las denominadas 7 herramientas clásicas para el control de la calidad constituyen un conjunto de instrumentos para la recopilación sistemática de datos y el análisis de resultados. Las 7 herramientas clásicas son las siguientes:

1. Hoja de verificación
2. Tormenta de ideas
3. Diagrama causa – efecto
4. Histograma
5. Diagrama de Pareto
6. Diagrama de dispersión
7. Gráfica de control

La primera actividad orientada al desarrollo en la mejora de la calidad antes de utilizar las 7 herramientas de la calidad es la colección de datos, pues los datos representan la información más importante que se obtiene acerca del comportamiento del proceso. Se cuenta principalmente con dos tipos de datos:

Datos variables. Son el resultado de las lecturas tomadas por aparatos de medición, son medibles y a su vez controlables; corresponden a datos tomados dentro de un rango de valores. Dependiendo de sus características, a dichas variables es posible determinarlas como: independientes, dependientes, discretas, continuas, dicotómicas, politómicas, etc.

Datos de atributos. Son medibles y solo se registra su ausencia o presencia, es decir, cuando se inspecciona un bien o servicio y se califica con pasa o no pasa.

Dentro de esta última categoría existen dos tipos de fallas:

- | | |
|---------------------|--|
| Defectos: | Son imperfecciones encontradas en un producto. |
| Defectuosos: | Se refiere al número de piezas rechazadas. |

Ahora bien, el recolectar estos datos requiere establecer monitoreo constante del proceso de producción, y se usan para llevar un análisis de lo que no se ajusta a las

especificaciones e inspeccionar con el fin de prevenir y mejorar la calidad del producto. Un método eficiente para recolectar datos puede ser el realizar una hoja de verificación, concepto que representa la primera de las siete herramientas de mejora y que se presenta a continuación.

1.3.1 Hoja de verificación

La hoja de verificación es un formato construido para coleccionar datos, de forma que su registro sea sencillo, sistemático y que sea fácil analizarlos. Una buena hoja de verificación debe reunir la característica de que, visualmente, permita hacer un primer análisis para apreciar las principales características de la información buscada. Algunas de las situaciones en las que resulta de utilidad obtener datos a través de las hojas de verificación son las siguientes:

- Describir el desempeño o los resultados de un proceso.
- Clasificar las fallas, quejas o defectos detectados, con el propósito de identificar sus magnitudes, razones, tipos de fallas, áreas de donde proceden, etcétera.
- Confirmar posibles causas de problemas de calidad.
- Analizar o verificar operaciones y evaluar el efecto de los planes de mejora.

La finalidad de la hoja de verificación es fortalecer el análisis y la medición del desempeño de los diferentes procesos de la empresa, con el fin de contar con información que permita orientar esfuerzos, actuar y decidir objetivamente.

En la tabla 1.3 se muestra un ejemplo de una hoja de verificación.

Hoja de Verificación					
Intervalos	Marcas de Clase	Tabulación	Frecuencia	Frecuencia Acumulada	Frecuencia Relativa
5,45 - 5,55	5,50	I	1	1	0,02
5,55 - 5,65	5,60	IIII II	7	8	0,14
5,65 - 5,75	5,70	IIII IIII I	11	19	0,22
5,75 - 5,85	5,80	IIII IIII IIII	15	34	0,30
5,85 - 5,95	5,90	IIII IIII	9	43	0,18
5,95 - 6,05	6,00	IIII	5	48	0,10
6,05 - 6,15	6,10	II	2	50	0,04
	TOTAL.		50		1,00
Realizó:			Departamento:		
Revisó:			Fecha:		

Tabla 1.3 Hoja de verificación. Elaboración propia

1.3.2 Tormenta de ideas

La tormenta de ideas es usada para identificar posibles soluciones a problemas y oportunidades potenciales de mejoramiento de la calidad. Además, es una técnica para

desarrollar el pensamiento creativo de grupo que genera y posteriormente clarifica una lista de ideas relacionadas con la solución de un problema.

La técnica de la tormenta de ideas se enfoca en la participación grupal organizada, conducente a la formación de ideas sobre una situación problemática. Haciendo la participación de cada uno de los miembros del grupo parte indispensable durante el proceso de tormenta de ideas, se tiene al final una visión global del problema y sus posibles causas.

1.3.3 Un diagrama causa – efecto

El diagrama causa – efecto es una herramienta utilizada para el desarrollo del pensamiento creativo y despliegue de las relaciones entre un efecto dado (variaciones en las características de calidad) y su causa potencial

Un diagrama causa – efecto tiene su fundamento en el análisis causal, el cual analiza las causas de los posibles efectos, con el fin de establecer posibles soluciones a dichas causas; además muestra la relación entre una característica determinada y los factores relacionados a esta, ayudando así a encontrar las causas posibles que la afectan, con lo cual es viable en una fase posterior determinar posibles soluciones a dichas causas.

En la figura 1.3 se muestra cómo se organiza el diagrama a partir de las posibles causas potenciales, que son organizadas en categorías mayores y sub–categorías formado de esta manera el despliegue de la problemática, similar a un esqueleto de pescado. De aquí que la herramienta también sea conocida como diagrama de espina de pescado.

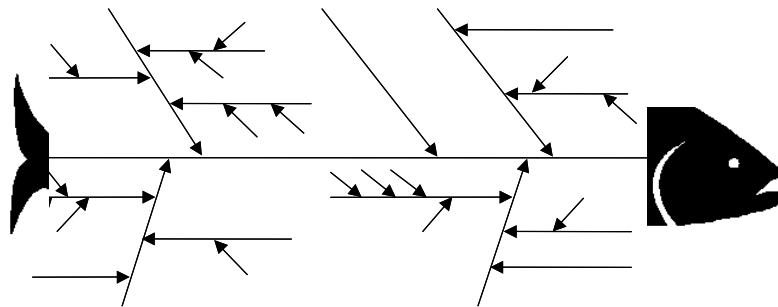


Figura 1.3 Esquema de diagrama causa-efecto. Elaboración propia

Un diagrama causa – efecto es utilizado para:

- Analizar y comunicar relaciones causa – efecto
- Facilitar la solución de problemas de un síntoma a una causa a una solución.

1.3.4 Histograma

El histograma es una representación gráfica, en forma de barras, de la distribución de un conjunto de datos o una variable, donde los datos se clasifican por su magnitud en cierto número de grupos o clases, y cada clase es representada por una barra, cuya altura es proporcional a la frecuencia de los valores representados. Por lo general, el eje horizontal está formado por una escala numérica para mostrar la magnitud de los datos, mientras que en el eje vertical se representan las frecuencias.

El histograma permite visualizar la tendencia central, la dispersión y la forma de la distribución. La figura 1.4 muestra seis histogramas con diseños de variación comúnmente encontrados. Examinando estos diseños, uno puede obtener pistas acerca del comportamiento de un proceso.

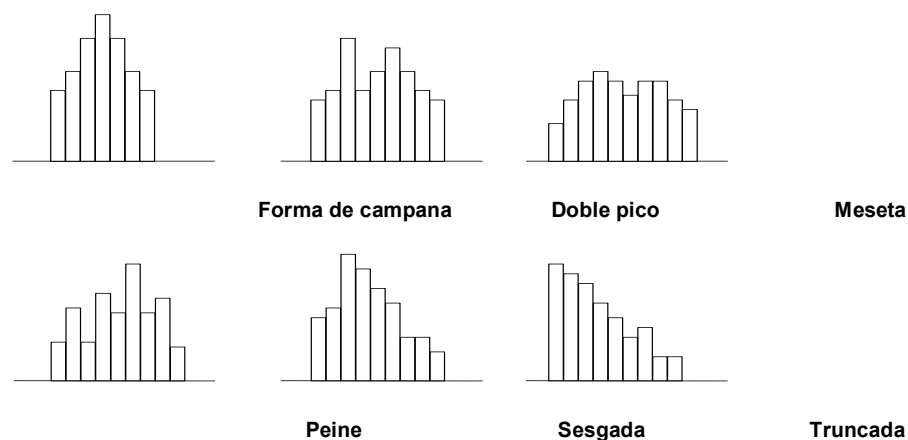


Figura 1.4 Tipos de distribución de frecuencias. Recuperado de Gutierrez (2009)

1.3.5 Diagrama de Pareto

El diagrama de Pareto es un gráfico especial de barras cuyo campo de análisis o aplicación son los datos categóricos, y tiene como objetivo ayudar a localizar el o los problemas vitales, así como sus principales causas. Generalmente, más del 80% de la problemática en una organización ocurre por causas comunes, es decir, se debe a problemas o situaciones que actúan de manera permanente sobre los procesos. Pero, además, en todo proceso son pocos los problemas o situaciones vitales que contribuyen en gran medida a la problemática global de un proceso o una empresa.

La idea es que cuando se quiere mejorar un proceso o atender sus problemas, no se den “palos de ciego” y se trabaje en todos los problemas al mismo tiempo atacando todas sus causas a la vez, sino que, con base en los datos e información aportados por un análisis estadístico, se establezcan prioridades y se enfoquen los esfuerzos donde éstos tengan mayor impacto. La utilidad general del diagrama está respaldada por el llamado *principio de Pareto*, conocido como “ley 80-20” o “pocos vitales, muchos triviales”, en el cual se reconoce que pocos elementos (20%) generan la mayor parte del efecto (80%), y el resto

de los elementos propician muy poco del efecto total. El nombre del principio se determinó en honor al economista italiano Wilfredo Pareto (1843-1923).

Para ejemplificar un poco más en el diagrama de Pareto, en la tabla 1.4 se presentan datos que comprenden una muestra de 500 quejas registradas sobre problemas en las líneas telefónicas.

Problema	Número de quejas	Porcentaje	Acumulado
Ruido en la línea	240	48 %	48 %
Línea abierta	120	24 %	72 %
Alarma	80	16 %	88 %
Sin respuesta	50	10 %	98 %
Sin sonido	10	2 %	100 %
Total	500	100 %	

Tabla 1.4 Quejas registradas. Elaboración propia

En la figura 1.5 se muestra un Diagrama de Pareto con base en la información mostrada en la tabla 1.4 mostrada anteriormente.

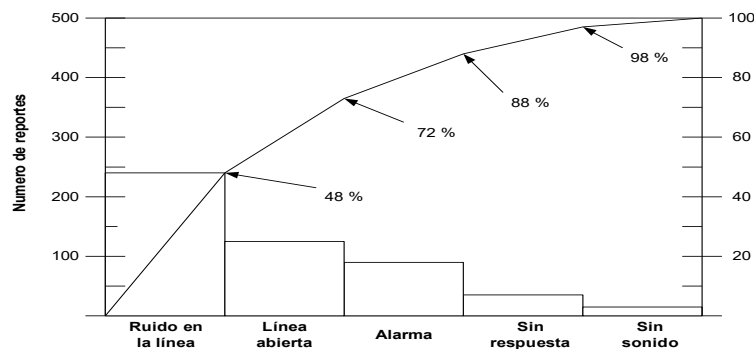


Figura 1.5 Diagrama de Pareto de quejas. Elaboración propia

1.3.6 Diagrama de dispersión

Dadas dos variables numéricas X y Y , medidas usualmente sobre el mismo elemento de la muestra de una población o proceso, el *diagrama de dispersión* es una gráfica del tipo X - Y , donde cada elemento de la muestra es representado mediante un par de valores (x_i, y_i) y el punto correspondiente en el plano cartesiano X - Y .

El objetivo de esta gráfica es analizar la forma en que estas dos variables están relacionadas; por ejemplo, estudiar en un grupo de estudiantes la relación entre su estatura (X) y su peso (Y). O podría ser de interés investigar la relación entre una variable de entrada (X) de un proceso con el valor de alguna característica de calidad (Y) del producto final. Al graficar todos los puntos, es decir, todas las parejas de valores (x_i, y_i) , si se observa que los puntos siguen algún patrón definido, esto será un indicio de una posible relación entre las dos variables.

El coeficiente de correlación es de utilidad para asegurarse de que la relación entre dos variables que se observa en un diagrama no se debe a una construcción errónea del diagrama de dispersión (por ejemplo, el tamaño y las escalas), así como para cuantificar la magnitud de la correlación lineal en términos numéricos. Para un conjunto de n valores del tipo (x_i, y_i) , obtenidos a partir de n unidades o productos, este coeficiente se calcula con:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} * S_{yy}}}$$

donde

$$S_{xy} = \sum_{i=1}^n (x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n})$$

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

$$S_{yy} = \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

Los valores que toma el coeficiente de correlación, r , están entre -1 y 1 , incluyendo $(-1 \leq r \leq 1)$. Los valores de r cercanos o iguales a cero implican poca o nula relación lineal entre X y Y . En contraste, los valores de r cercanos a 1 indican una relación lineal muy fuerte, y los valores de r próximos a -1 muestran una fuerte correlación negativa. Los valores de r menores a -0.85 o mayores a 0.85 indican una correlación fuerte; mientras que los valores de r entre -0.50 y 0.50 se refieren a una correlación de moderada a débil. Por último, los valores de r iguales o menores que -0.30 o 0.30 indican una correlación lineal prácticamente inexistente. Cabe mencionar que, aun cuando existe una fuerte relación lineal entre dos variables, esta relación no necesariamente es causal.

En la figura 1.6 se muestran algunos patrones de correlación que se pueden presentar en un diagrama de dispersión.

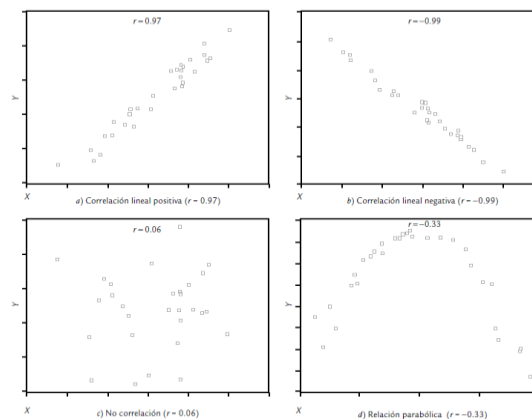


Figura 1.6 Patrones de correlación. Recuperado de Gutierrez (2009)

1.3.7 Gráfica de control

Una gráfica o diagrama de control es una herramienta para detectar variaciones debido a causas especiales o asignables a situaciones indeseables en un proceso. Las variaciones inherentes al proceso generalmente se repiten aleatoriamente dentro de ciertos límites preestablecidos (límites de control); variaciones atribuibles generan puntos fuera de estos límites de control. Estas variaciones debido a causas asignables o especiales indican que algunos de los factores que afectan al proceso se deben identificar, investigar y llevar a control.

La construcción de gráficos de control tiene su fundamento en la estadística. Los gráficos de control utilizan datos operativos para el establecimiento de límites, dentro de los cuales observaciones futuras son llevadas a cabo, para determinar si el proceso permanece estable sin la presencia de causas especiales o asignables de variación.

Para más detalle sobre las gráficas de control véase el punto 1.5. Asimismo, además de las 7 herramientas básicas de control propuestas por Kaoru Ishikawa, hay otras herramientas estadísticas de interés dentro del estudio de la calidad, como son el muestreo, los estudios de capacidad del proceso, la prueba de hipótesis y el muestreo de aceptación.

1.4 Muestreo de datos

El muestreo es el estudio de una selección de elementos de una población, llamada muestra, con el fin de estimar los parámetros de la población, mediante la inferencia estadística.

En la industria es importante contar con información sobre el comportamiento del proceso para determinar qué está sucediendo en el mismo; sin embargo, analizar toda la población tendría un costo asociado elevado, por lo que mejor será examinar sólo una parte de ella mediante un muestreo. Trabajar con una muestra de la población tiene la ventaja de que es más rápido, más barato y los resultados obtenidos pueden ser casi tan precisos como cuando se analiza toda la población, si la estadística se usa adecuadamente: si la muestra se elige correctamente, la información que se obtiene permite una estimación razonable de la situación de la población.

Existen diferentes tipos de muestro que se mencionarán a continuación.

1. **Muestreo aleatorio simple.** Este consiste en seleccionar los datos de manera que todos tengan la misma probabilidad de formar parte de la muestra.
2. **Muestreo aleatorio sistemático.** En este caso se elige el primer individuo al azar y el resto viene condicionado por aquél. Puede aplicarse en la mayoría de las situaciones, la única precaución que debe tenerse en cuenta es comprobar que la característica que se estudia no tenga una periodicidad que coincida con la del muestreo. Es útil cuando se busca seleccionar individuos a lo largo del tiempo.

Por ejemplo, para estudiar la satisfacción de un servicio, se puede elegir sistemáticamente encuestar a 1 de cada n clientes que visitan la sucursal. En estas circunstancias, en las que puede existir diferente varianza entre individuos en diferentes periodos de tiempo, el muestreo sistemático puede ser incluso más preciso que el muestreo aleatorio simple.

3. **Muestreo aleatorio estratificado.** Se divide la población en grupos en función de un carácter determinado y después se muestrea cada grupo aleatoriamente, para obtener la parte proporcional de la muestra. Este método se aplica para evitar que por azar algún grupo de datos este menos representado que los otros. Además de tener resultados para el conjunto completo, se tiene información también de los subgrupos (estratos).
4. **Muestreo aleatorio por conglomerados.** Se divide la población en varios grupos de características parecidas entre ellos y luego se analizan completamente algunos de los grupos, descartando los demás. Dentro de cada conglomerado existe una variación importante, pero los distintos conglomerados son parecidos. Requiere una muestra más grande, pero suele simplificar la recogida de muestras y por tanto generalmente es más barato. Frecuentemente los conglomerados se aplican a zonas geográficas.

1.5 Prueba de hipótesis

En inferencia estadística, con un conjunto de datos muestrales, se plantea inferir sobre la población. Esas inferencias o decisiones que se pueden tomar son referentes a estimación de parámetros o prueba de hipótesis. Al procedimiento mediante el cual se investiga la verdad o falsedad de una hipótesis se le llama prueba de hipótesis. Cabe destacar que en distintas disciplinas del conocimiento se requiere que se tome una decisión entre aceptar o rechazar una proposición sobre algún parámetro.

Dentro del control estadístico de la calidad existen muchas razones para considerar las pruebas de hipótesis como un instrumento indispensable; entre algunas de las que se pueden mencionar están las siguientes:

- a) Son las herramientas de trabajo para un buen análisis, esto es, de la información de los procesos se pueden deducir hipótesis.
- b) De estas hipótesis se puede establecer que son probablemente ciertas o probablemente falsas.
- c) Son un instrumento poderoso, porque ayudan a confirmar o negar una suposición en forma independiente de la opinión del especialista que esté desarrollando el análisis.

Debe hacerse hincapié en que la verdad o falsedad de una hipótesis en particular nunca puede conocerse con certidumbre, a menos de que pueda examinarse a toda la población. Esto es imposible en muchas situaciones prácticas. Por lo tanto, es necesario

desarrollar un procedimiento de prueba de hipótesis teniendo en cuenta la probabilidad de llegar a una conclusión equivocada.

La hipótesis nula, representada por H_0 es la afirmación sobre una o más características de poblaciones que al inicio se supone cierta (es decir, la "creencia a priori").

La hipótesis alternativa, representada por H_a , es la afirmación contradictoria a H_0 , y ésta es un enunciado sobre la población. Su posibilidad de ser evaluada es con base en la información obtenida de una muestra aleatoria de la población.

Con base en lo anterior, la hipótesis nula se caracterizó como la afirmación hecha sobre la población; tal afirmación puede tener dos resultados posibles y complementarios al probar su validez, estos son:

La hipótesis H_0 es cierta.

La hipótesis H_0 es falsa.

Si los datos sugieren que la hipótesis nula es falsa, entonces se debe rechazar y concluir: la hipótesis H_0 no está ampliamente respaldada por los datos de la muestra; en caso contrario se concluirá que la hipótesis H_0 está ampliamente respaldada por los datos de la muestra.

En una prueba de hipótesis, la toma de decisiones se ha sujeta a los datos recolectados a través de un experimento o de una muestra aleatoria, por lo que es posible cometer dos tipos de errores que pueden llevar a una pérdida sustancialmente diferente, estos errores son:

- Error tipo I: Rechazar una hipótesis verdadera
- Error tipo II: No rechazar una hipótesis falsa

En la tabla 1.5 se muestra esquemáticamente este tipo de errores que se pueden cometer relacionando las decisiones que se pueden tomar de acuerdo con la situación real del problema.

		Situación real de H_0	
		<i>Cierta</i>	<i>Falsa</i>
Decisión	Rechazar H_0	Error de tipo I α	Decisión
	No rechazar H_0	Decisión correcta	Error de tipo II β

Tabla 1.5 Reglas de decisión de una hipótesis estadística. Elaboración propia

El tamaño de estos dos tipos de error se define como la probabilidad de que cada uno de ellos ocurra; al término α (alfa) se le llama nivel de significancia de la prueba, generalmente se le asignan los valores 0.10, 0.05 ó 0.01 y se determina al inicio de la investigación; a β (beta) se le denomina potencia de la prueba. Nótese que, debido a la relación existente entre las hipótesis, los dos tipos de error se hallan relacionados, y al

controlar el error tipo I automáticamente se controla el otro para un tamaño de muestra dado.

En la práctica se pueden presentar tres diferentes tipos de prueba. Sea θ un parámetro poblacional y θ_0 un valor que se obtiene en un experimento, el cual se comparará con el parámetro de la población a través de una hipótesis. Las diferentes formas que puede presentar una prueba estadística de hipótesis son las siguientes:

Unilateral derecha	Unilateral izquierda	Bilateral
$H_0: \theta = \theta_0$	$H_0: \theta = \theta_0$	$H_0: \theta = \theta_0$
$H_a: \theta > \theta_0$	$H_a: \theta < \theta_0$	$H_a: \theta \neq \theta_0$

1.6 Estudio de capacidad del proceso

Los procesos tienen variables de salida o de respuesta, las cuales deben cumplir con ciertas especificaciones para saber que el proceso está funcionando de manera satisfactoria. Evaluar la habilidad o capacidad de un proceso consiste en conocer la amplitud de la variación natural de éste para una característica de calidad dada, lo cual permitirá saber en qué medida tal característica de calidad cumple las especificaciones.

Esta variación es denominada por causas comunes (o por azar) es aquella que permanece día a día, lote a lote; y es aportada de forma natural por las condiciones de las 6 M (materiales, maquinaria, medición, mano de obra, métodos y medio ambiente). Esta variación es resultado de la acumulación y combinación de diferentes causas que son difíciles de identificar y eliminar, ya que son inherentes al sistema y la contribución individual de cada causa es pequeña; no obstante, a largo plazo representan la mayor oportunidad de mejora.

Un proceso es capaz cuando se asegura que casi la totalidad de la producción cumplirá con las especificaciones establecidas por el cliente.

Para relacionar la variabilidad del proceso con los límites especificados, se definen los siguientes índices de capacidad sus definiciones y formulas fueron consultados en (Gutierrez, 2009).

- **Índice Cp.** El índice de capacidad potencial del proceso, C_p , se define como el cociente entre la variación tolerada del proceso y la variación real observada del mismo.

$$C_p = \frac{LSE - LIE}{6\sigma_x}$$

En donde:

- LSE: límite superior de especificación
- LIE: límite inferior de especificación
- $6\sigma_x$ corresponde al rango de tolerancia natural del proceso y contiene el 99.73% de los productos.

Interpretación del índice C_p . Para que el proceso sea considerado potencialmente capaz de cumplir con especificaciones, se requiere que la variación real (natural) siempre sea menor que la variación tolerada. En la tabla 1.6 se muestra los valores de C_p para diferentes categorías de proceso, según sea la habilidad que tenga para cumplir con las especificaciones.

Valor de C_p (corto plazo)	Categoría de proceso	Decisión (si el proceso está centrado)
$C_p \geq 2$	Clase mundial	Se tiene calidad Seis Sigma.
$C_p \geq 1.33$	1	Adecuado.
$1 < C_p < 1.33$	2	Parcialmente adecuado. Requiere de un control estricto
$0.67 < C_p < 1$	3	No adecuado. Se requiere análisis y adaptación del proceso para alcanzar una calidad satisfactoria.
$C_p < 0.67$	4	No adecuado. Requiere de modificaciones serias.

Tabla 1.6 Clasificación de los procesos de acuerdo con su capacidad. Recuperado de Gutierrez (2009)

La capacidad potencial, C_p , no es el único índice que se utiliza para determinar la capacidad de un proceso industrial: tiene la desventaja de no tomar en cuenta el grado de centrado del proceso. Se han desarrollado otros índices, algunos de los cuales se presentan a continuación.

- **Índice C_r .** Un índice menos conocido que el C_p , es el que se conoce como razón de capacidad potencial, C_r , y está definido por:

$$C_r = \frac{6\sigma_x}{LSE - LIE}$$

Como se puede apreciar, el índice C_r es el inverso del C_p , ya que compara la variación real frente a la variación tolerada. La ventaja del índice C_r sobre el C_p es que su interpretación es un poco más intuitiva, a saber: el valor del índice C_r representa la proporción de la banda de especificaciones que es ocupada por el proceso.

- **Índices laterales C_{pi} , C_{ps} y la capacidad real C_{pk} .** La desventaja de los índices C_p y C_r es que no toman en cuenta el centrado del proceso, debido a que en las fórmulas para calcularlos no se incluye de ninguna manera la media del proceso,

μ_x Una forma de corregir esto consiste en evaluar por separado el cumplimiento de la especificación inferior y superior, a través del índice de capacidad para la especificación inferior, C_{pi} , y índice de capacidad para la especificación superior, C_{ps} , respectivamente, los cuales se calculan de la siguiente manera:

$$C_{pi} = \frac{\mu_x - LIE}{3\sigma_x}$$

$$C_{ps} = \frac{LSE - \mu_x}{3\sigma_x}$$

Por su parte, el índice C_{pk} , que se conoce como índice de capacidad real del proceso, es considerado una versión corregida del C_p que sí toma en cuenta el centrado del proceso.

$$C_{pk} = \min\{C_{pi}, C_{ps}\}$$

Si $C_{pk} < 1$, entonces el proceso no cumple con por lo menos una de las especificaciones. Algunos elementos adicionales para la interpretación del índice C_{pk} son los siguientes:

- El índice C_{pk} siempre va a ser menor o igual que el índice C_p cuando son muy próximos, eso indica que la media del proceso está muy cerca del punto medio de las especificaciones, por lo que la capacidad potencial y real son similares.
 - Si el valor del índice C_{pk} es mucho más pequeño que el C_p , significa que la media del proceso está alejada del centro de las especificaciones. De esa manera, el índice C_{pk} estará indicando la capacidad real del proceso, y si se corrige el problema de descentrado se alcanzará la capacidad potencial indicada por el índice C_p .
 - Cuando el valor del índice C_{pk} sea mayor a 1.25 en un proceso ya existente, se considerará que se tiene un proceso con capacidad satisfactoria. Mientras que para procesos nuevos se pide que $C_{pk} > 1.45$.
 - Es posible tener valores del índice C_{pk} iguales a cero o negativos, e indican que la media del proceso está fuera de las especificaciones.
- **Índice de localización, K .** El índice de descentrado de proceso o índice de localización, K , mide en términos relativos y porcentuales qué tan descentrado está el proceso con respecto al valor nominal N :

$$K = \frac{\mu_x - N}{\frac{1}{2}(LSE - LIE)} * 100\%$$

El valor nominal, N, corresponde a la calidad objetiva y óptima. Cualquier desviación de este valor significa una disminución en la calidad, ya que el descentrado del proceso provoca la producción de unidades fuera de especificación.

- Si el signo del valor de K es positivo significa que la media del proceso es mayor al valor nominal y será negativo cuando $\mu < N$.
- Valores de K menores a 20% en términos absolutos se consideran aceptables, pero a medida que el valor absoluto de K sea más grande que 20%, indica un proceso muy descentrado, lo cual contribuye de manera significativa a que la capacidad del proceso para cumplir especificaciones sea baja.
- **Índice C_{pm}** (índice de Taguchi) Los índices C_p y C_{pk} están pensados a partir de lo importante que es reducir la variabilidad de un proceso para cumplir con las especificaciones. Sin embargo, desde el punto de vista de G. Taguchi, cumplir con especificaciones no es sinónimo de buena calidad y la reducción de la variabilidad debe darse en torno al valor nominal (calidad óptima), en consecuencia, Taguchi (1986) propone que la capacidad del proceso se mida con el índice C_{pm} que está definido por:

$$C_{pm} = \frac{LSE - LIE}{6\tau}$$

Donde: $\tau = \sqrt{\sigma_x^2 + (\mu_x - N)^2}$

- **Índice Z.** Otra forma de medir la capacidad del proceso es mediante el índice Z, el cual consiste en calcular la distancia entre las especificaciones y la media μ del proceso en unidades de la desviación estándar, σ . De esta manera, para un proceso con doble especificación se tiene z superior, z_s , y z inferior, z_i , que se definen de la siguiente manera:

$$z_i = \frac{\mu_x - LIE}{\sigma_x}$$

$$z_s = \frac{LSE - \mu_x}{\sigma_x}$$

1.7 Diagramas de control

Varios autores presentan explicaciones claras sobre lo que es un diagrama de control. La información presentada en este inciso está basada en los libros de Gutiérrez (2009) y Breyfogle (2003).

Esta teoría general de los gráficos de control fue propuesta por primera vez por Shewhart en 1924. La publicación en 1931 del libro de Shewhart, *Economic control of quality of manufactured products*, marcó el inicio de control estadístico de procesos.

Las gráficas de control son gráficas poligonales que muestran el estado del proceso en el tiempo; son elementos indispensables en manos de quienes deben resolver problemas de cambios en el proceso por causas específicas, porque proporcionan información sobre:

- El intervalo de variación en el que básicamente se mueve la característica de calidad.
- La consistencia del proceso de producción en el tiempo.
- El nivel medio de la característica de calidad, cuyo conocimiento es básico para la toma de decisiones.

Las gráficas de control se usan, entre otras cosas, para:

- verificar que los datos obtenidos vienen de procesos con condiciones semejantes.
- observar el proceso productivo, a fin de poder investigar las causas de un comportamiento anormal.

Existen diferentes gráficas de control en función de la variable a observar y del proceso a controlar. El proceso por controlar puede basarse en el registro de una variable o un atributo. Las gráficas de control tradicionales más utilizadas son las siguientes:

Gráficas de control para variables:

- \bar{X} (de medias muestrales)
- R (de rangos)
- S (de desviaciones estándar)
- X (de medidas individuales)

Gráficas de control para atributos:

- p (proporción o fracción de artículos defectuosos)
- np (número de unidades defectuosas)
- c (número de defectos)
- u (número de defectos por unidad)

Los gráficos de control son la herramienta más poderosa del control estadístico de procesos. Su finalidad es conseguir y mantener un proceso bajo control estadístico mediante la reducción sistemática de la variabilidad atribuible a causas específicas.

Esta variación es denominada por causas especiales (o atribuibles) es causada por situaciones o circunstancias especiales que no están de manera permanente en el proceso. Por ejemplo, la falla ocasionada por el mal funcionamiento de una pieza de la máquina, el empleo de materiales no habituales o el descuido de un operario. Las causas especiales, por su naturaleza relativamente discreta, a menudo pueden ser identificadas y eliminadas si se cuenta con los conocimientos y condiciones para ello.

Cuando un proceso trabaja sólo con causas comunes de variación se dice que está bajo control estadístico o es estable, porque su variación a través del tiempo es predecible. Además, independientemente de que su variabilidad sea mucha o poca, el desempeño del proceso es predecible en el futuro inmediato, en el sentido de que su tendencia central y la amplitud de su variación se mantienen sin cambios, al menos en el corto plazo. En contraste, se dice que un proceso en el que están presentes causas especiales de variación está fuera de control estadístico (o simplemente que es inestable); este tipo de procesos son impredecibles en el futuro inmediato pues en cualquier momento pueden aparecer de nuevo las situaciones que tienen un efecto especial sobre la tendencia central o sobre la variabilidad. No distinguir entre estos dos tipos de variabilidad conduce a cometer dos errores en la actuación de los procesos.

Una de las tantas preguntas es saber qué gráfico de control seleccionar con base en el tipo y número de datos, por lo cual en la figura 1.7 se presenta un diagrama de flujo que sirve de ayuda para saber qué tipo de gráfico emplear.

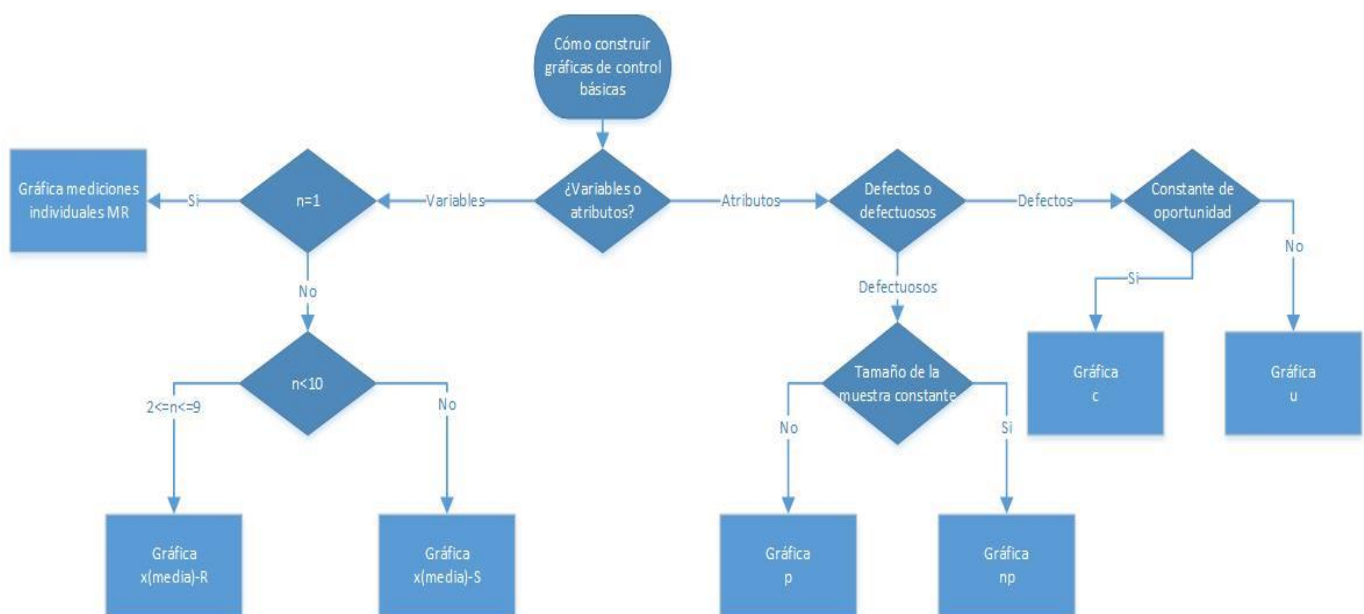


Figura 1.7 Diagrama de flujo para selección de gráficos de control. Elaboración propia

De acuerdo con el diagrama de flujo anterior se selecciona el tipo apropiado de gráfico de control y se calculan los estadísticos, así como los límites de control basado en los estadísticos del muestreo de los elementos muestreados en cada momento en el tiempo. Cabe mencionar que también existen graficas de control para detectar cambios pequeños en el proceso como son CUSUM y EWMA, que se detallaran más adelante.

Las gráficas de control para variables se aplican a características de calidad de tipo continuo, que intuitivamente son aquellas que se miden con un instrumento de medición (peso, volumen, voltaje, longitud, resistencia, temperatura, humedad, etc.). La gráfica de control en realidad corresponde a una secuencia de pruebas de hipótesis de que la característica de control no haya cambiado en el tiempo; los límites de control corresponden a los límites de aceptación de la prueba de hipótesis anterior. En la tabla 1.7 se muestra información de cómo calcular los límites de control dependiendo de qué tipo de gráfico se quiera construir; en este caso es para diagramas de control para variables.

Tipo de gráfico	LIC	Línea central	LSC
$\bar{X} - R$	$\bar{\bar{X}} - 3\left(\frac{\bar{R}/d_2}{\sqrt{n}}\right)$	$\bar{\bar{X}}$	$\bar{\bar{X}} + 3\left(\frac{\bar{R}/d_2}{\sqrt{n}}\right)$
$\bar{X} - S$	$\bar{\bar{X}} - 3\left(\frac{\bar{S}}{c_4\sqrt{n}}\right)$	$\bar{\bar{X}}$	$\bar{\bar{X}} + 3\left(\frac{\bar{S}}{c_4\sqrt{n}}\right)$
$\bar{X} - R$ (Medidas individuales)	$\bar{\bar{X}} - 3\frac{\bar{R}}{1.128}$	$\bar{\bar{X}}$	$\bar{\bar{X}} + 3\frac{\bar{R}}{1.128}$
R (Rangos)	$\bar{\bar{R}} - 3d_3\left(\frac{\bar{R}}{d_2}\right)$	$\bar{\bar{R}}$	$\bar{\bar{R}} + 3d_3\left(\frac{\bar{R}}{d_2}\right)$
S (Desviación estándar)	$\bar{\bar{S}} - 3\frac{\bar{S}}{c_4}\sqrt{1 - c_4^2}$	$\bar{\bar{S}}$	$\bar{\bar{S}} + 3\frac{\bar{S}}{c_4}\sqrt{1 - c_4^2}$

Tabla 1.7 Fórmulas para el cálculo de límites de control para variables. Elaboración propia

Si el producto no tiene la calidad deseada, no se permite que pase a la siguiente etapa del proceso: es segregado y se le denomina artículo defectuoso.

En la tabla 1.8 se muestra de igual manera información de cómo calcular los límites de control dependiendo de qué tipo de gráfico se quiera construir; en este caso es para diagramas de control para atributos.

Tipo de gráfico	LIC	Línea central	LSC
p (Proporción de defectuosos)	$\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	\bar{p}	$\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$
np (Número de defectuosos)	$n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$	$n\bar{p}$	$n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$
c (Número de defectos)	$\bar{c} - 3\sqrt{\bar{c}}$	\bar{c}	$\bar{c} + 3\sqrt{\bar{c}}$
u (Número de defectos por unidad)	$\bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$	\bar{u}	$\bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$

Tabla 1.8 Fórmulas para el cálculo de límites de control para atributos. Elaboración propia

Una desventaja de los gráficos de control tradicionales (tipo Shewhart), es que no son rápidas para detectar “cambios pequeños” en el proceso. Para solucionar este problema existen las gráficas tipo CUSUM y EWMA.

- **CUSUM.** Este gráfico fue propuesto por Page (1954), y el nombre de CUSUM se debe a que es un gráfico que construye a través de la suma acumulada de las desviaciones con respecto a la media global (si el proceso está centrado, se podrían considerar las desviaciones con respecto al valor nominal de la característica de interés). Entonces, en los primeros m puntos de inspección sobre el diagrama CUSUM se grafican las sumas acumuladas:

$$S_1 = (\bar{X}_1 - \hat{u})$$

$$S_2 = (\bar{X}_1 - \hat{u}) + (\bar{X}_2 - \hat{u})$$

$$S_3 = (\bar{X}_1 - \hat{u}) + (\bar{X}_2 - \hat{u}) + (\bar{X}_3 - \hat{u})$$

.

.

.

$$S_m = (\bar{X}_1 - \hat{u}) + (\bar{X}_2 - \hat{u}) + (\bar{X}_3 - \hat{u}) + \dots + (\bar{X}_m - \hat{u}) = \sum_{i=1}^m (\bar{X}_i - \hat{u})$$

- **EWMA.** La gráfica EWMA (por sus siglas en inglés: Exponentially Weighted Moving-Average, “promedios móviles exponencialmente ponderados” fue propuesta por Roberts (1959). Este gráfico tiene un desempeño muy parecido a la CUSUM en la detección de pequeños cambios de nivel del proceso. El estadístico EWMA, que se grafica al tiempo t en la gráfica, está dado por la siguiente fórmula:

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}$$

donde $Z_0 = \bar{\bar{X}}$ que coincide con el valor nominal si el proceso está centrado, y $0 < \lambda \leq 1$. Por ejemplo, hasta Z_3 , las tres primeras sumas estarían dadas por:

$$Z_1 = \lambda \bar{x}_1 + (1 - \lambda)Z_0$$

$$Z_2 = \lambda \bar{x}_2 + (1 - \lambda)Z_1$$

$$Z_3 = \lambda \bar{x}_3 + (1 - \lambda)Z_2$$

A partir de esto, el parámetro λ determina la profundidad de la memoria de la EWMA: mientras más cerca esté de cero es mayor el peso de los datos históricos, es decir, recuerda más el pasado. Por otro lado, si está más cerca de uno, tiene más influencia la última media observada y el pasado tiene menos peso. Cuando $\lambda = 1$, el diagrama de control será equivalente al gráfico de promedios tradicional, que no da ningún peso a la información anterior a un punto dado. La experiencia ha mostrado que lo adecuado es que $0.1 \leq \lambda \leq 0.3$, y el valor 0.2 es el más típico.

La varianza del estadístico Z_t está dada por:

$$Var(Z_t) = \frac{\sigma^2}{n} \left[\frac{\lambda}{2 - \lambda} \right] [1 - (1 - \lambda)^{2t}]$$

donde n es el tamaño del subgrupo. De aquí que los límites en el punto o subgrupo t están dados por:

$$LCS = Z_0 + 3 \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\left[\frac{\lambda}{2 - \lambda} \right] [1 - (1 - \lambda)^{2t}]}$$

$$LCI = Z_0 - \frac{3\hat{\sigma}}{\sqrt{n}} \sqrt{\left[\frac{\lambda}{(2 - \lambda)} \right] [1 - (1 - \lambda)^{2t}]}$$

Como en las expresiones anteriores, el término $[1 - (1 - \lambda)^{2t}]$ tiende a 1 cuando t se incrementa, de manera que la varianza de Z_t se va incrementando y con ello los límites de control del gráfico EWMA se van abriendo en los primeros puntos hasta estabilizarse en:

$$LCS = Z_0 + \frac{3\hat{\sigma}}{\sqrt{n}} \sqrt{\left[\frac{\lambda}{(2 - \lambda)} \right]}$$

$$LCI = Z_0 - \frac{3\hat{\sigma}}{\sqrt{n}} \sqrt{\left[\frac{\lambda}{(2 - \lambda)} \right]}$$

1.8 Muestro de aceptación

En las actividades de control de calidad, en ocasiones es necesario inspeccionar lotes de materia prima, así como partes o productos terminados para asegurar que se cumplen ciertos niveles de calidad con un buen grado de confianza. El muestreo de aceptación es el proceso de inspección de una muestra de unidades extraídas de un lote que se realiza con el propósito de aceptar o rechazar todo el lote. Se puede aplicar en cualquier relación cliente-proveedor.

Se debe tener claro que el muestreo de aceptación es una forma particular de inspección, en la que simplemente se aceptan y rechazan lotes, pero no mejora la calidad. Es decir, este muestreo no es una estrategia de mejora de la calidad, sino más bien una estrategia para proporcionar cierto nivel de seguridad de que los niveles de calidad declarados para cierto producto se están alcanzando. Por lo tanto, es una estrategia defensiva ante el posible deterioro de la calidad.

De acuerdo con Montgomery (2016), hay tres enfoques para la dictaminación de lotes y son los siguientes:

- La aceptación sin inspección
- La inspección al 100%
- El muestreo de aceptación

Montgomery habla de ventajas y desventajas del muestro cuando se compara con la inspección al 100% y éstas son:

- Suele tener costos más bajos, debido a que hay menos inspección
- Hay menos manejo del producto y, en consecuencia, se reducen los daños
- Puede aplicarse en pruebas destructivas
- En las actividades de inspección participa menos personal
- La cantidad de errores de inspección generalmente se reduce en gran medida
- El rechazo de lotes completos, por oposición a la simple devolución de las unidades defectuosas, con frecuencia proporciona una motivación mayor para que el proveedor atienda el mejoramiento de calidad

El muestro de aceptación también presenta varias desventajas. Entre ellas se encuentran:

- Existe el riesgo de aceptar lotes “malos” y de rechazar lotes “buenos”
- Por lo general se genera menos información acerca del producto o acerca del proceso con que se fabricó el producto, que con un muestreo al 100%
- El muestreo de aceptación requiere la planeación y documentación del procedimiento de muestreo de aceptación, mientras que la inspección de todo el lote no lo requiere

El muestreo de aceptación es un “terreno intermedio” entre los extremos de la inspección del 100% y no hacer ninguna inspección.

1.9 ¿Qué es R?

R es un lenguaje de programación para la manipulación de datos, cálculo y representación gráfica. Este software ofrece una amplia variedad de técnicas estadísticas, por mencionar algunas: modelos de regresión lineal, pruebas estadísticas, análisis de series temporales, clasificación y agrupación de datos, etc.

En 1993, R fue desarrollado como lenguaje de programación por Ross Ihaka y Robert Gentleman, profesores de estadística de la Universidad de Auckland en Nueva Zelanda como un dialecto del lenguaje S, que había sido creado a principios de los años 90 del Siglo XX por los Laboratorios AT&T Bell (Jose Miguel Contreras Garcia, 2010).

En junio de 1995 ya se disponía de una versión pública de R para el cual el código fuente estuvo disponible en los términos de la licencia general GNU de la Free Software Foundation. Con este lanzamiento surgió una amplia comunidad de usuarios, como estadísticos y expertos en computación que, de manera desinteresada, comenzaron a desarrollar y publicar cambios en el código de R, los cuales eran comunicados a los fundadores a través de correo. De vez en cuando se liberaban versiones actualizadas; R se convirtió en un software libre debido a que los usuarios tienen la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar el software. El software libre es una cuestión de libertad, no de precio (GNU, 2016).

Como se menciona en la página de GNU, un software es libre si cumple con las cuatro libertades

- Libertad 0. La libertad de ejecutar el programa con cualquier propósito y como el usuario lo desee.
- Libertad 1. La libertad de estudiar el funcionamiento del programa y adaptarlo a sus necesidades. El acceso al código fuente es un prerrequisito para esto.
- Libertad 2. La libertad de redistribuir copias para ayudar a los demás.
- Libertad 3. La libertad de mejorar el programa y de publicar las mejoras, de modo que toda la comunidad se beneficie. El acceso al código fuente es un prerrequisito para esto.

En marzo de 1996, R tuvo un mayor desarrollo y la gente comenzó a portar más aplicaciones sobre el mismo, por lo cual para Ross y Robert era casi imposible darles seguimiento a todas estas observaciones por medio del correo. Tras un breve debate, se tomó la decisión de establecer el archivo en un sitio en Austria. Con esta implementación se obtuvieron más informes y sugerencias. La participación de la gente fue tanta que no siempre se podía hacer los cambios a una velocidad tal que esta fuera satisfactoria para los que pedían cambios.

A mediados de 1997, se estableció un "grupo central" más grande, eran personas que podían realizar cambios en el archivo de código fuente, dado que todo el trabajo en R es estrictamente de carácter voluntario, la organización tenía poco avance, porque los miembros contribuían cuando y como podían.

Para el año 2000 aparece la primera versión oficial para el público en general (la versión 1.0.0), desde entonces con el paso de los años se ha tenido un progreso en su desarrollo y actualmente está disponible la versión 3.5.1.

1.9.1 Características

R se ha convertido en un software muy potente a nivel internacional, al grado que, a estadísticos, ingenieros, científicos e investigadores, aún sin conocimientos previos de programación, les resulta fácil de usar. R permite hacer análisis desde muy básicos a complejos, dependiendo del propósito que cada usuario quiera lograr.

Es importante mencionar que R tiene cerca de 1,600 paquetes complementarios realizados por distintas personas alrededor del mundo; estos paquetes, antes de ser publicados en el sitio oficial correspondiente, pasan por una revisión por especialistas en el tema y una vez concluida ésta se publica para disposición del público en general. El número de paquetes ha crecido de manera exponencial.

Como resultado del trabajo comunitario que se ha realizado a través de los años, actualmente en R se puede realizar infinidad de tareas con base en sus características tales como:

- Un manejo eficaz de los datos
- Usar un conjunto de operadores para cálculos en matrices
- Una programación que incluye condicionales, bucles, funciones recursivas definidos por el usuario y las instalaciones de entrada y salida
- Realizar avanzadas operaciones de análisis de datos, que se presentan en pantalla y permiten obtener resultados intermedios (valores p , coeficientes de regresión, residuales, entre otros)
- Exportar la información a un archivo de datos para su visualización o para ser utilizados en posteriores análisis
- Importar de datos, por ejemplo, de CSV, SAS y SPSS, o directamente desde Microsoft Excel, Microsoft Access, Oracle, MySQL y SQLite
- Obtener representaciones gráficas de los análisis, que pueden ser visualizados en la propia ventana de R o ser exportados a distintos formatos (jpg, png, bmp, pdf, emf, pictex, xfig, entre otros)
- R puede ser extendido fácilmente a través de paquetes
- R tiene su propio formato de documentación similar al \LaTeX , que se utiliza para suministrar una amplia documentación, tanto en línea como en varios formatos

1.9.2 Descarga e instalación de R

1. Entrar en la página web: <http://www.r-project.org/>
2. Pulsar el conector *CRAN* bajo la referencia *Download*. Aparecen un conjunto de direcciones web donde está disponibles copias del software para ser descargadas. Por proximidad se puede elegir cualquiera de las siguientes tres opciones que aparecen en el apartado de México.

México

<https://cran.itam.mx/>

Instituto Tecnológico Autónomo de México

<http://cran.itam.mx/>

Instituto Tecnológico Autónomo de México

<http://www.est.colpos.mx/R-mirror/> Colegio de Postgraduados, Texcoco

3. En el recuadro *Download and Install R*, seleccionar el sistema operativo de interés (*Linux*, *MAC*, *Windows*)
4. Para descargar el paquete básico, dar click en el hipervínculo llamado *base* Resulta una página con el título *R-3.5.1 for Windows*. En ella, elegir el hipervínculo *Download R-3.5.1*, con lo que saldrá la siguiente ventana de diálogo
5. Se elige *Ejecutar*, es el programa ejecutable (.exe) instalador del sistema R básico. En la ventana de diálogo que resulta se elige *Ejecutar*. Posteriormente se solicita el idioma de instalación se selecciona el que sea de su preferencia y seguir los pasos que se indican en cada ventana hasta concluir con la instalación.

1.9.3 Rstudio

Rstudio es un entorno de desarrollo integrado (Integrated Development Environment, IDE, por sus siglas en inglés) para el lenguaje de programación R. De acuerdo con Venzani (2011). Un IDE es una aplicación de software que proporciona un entorno de programación para agilizar el desarrollo y la depuración de software.

Generalmente, un IDE proporciona una interfaz fácil de usar, automatiza los pasos de desarrollo y permite a los desarrolladores ejecutar y depurar programas desde una sola pantalla. También puede proporcionar el enlace desde un sistema operativo de desarrollo a una plataforma de destino de aplicaciones, como un entorno de escritorio, un teléfono inteligente o un microprocesador. En este sentido, Rstudio es un IDE de R.

Rstudio se inicia como la mayoría de las demás aplicaciones. En la figura 1.8 se puede ver la aplicación ejecutándose bajo una versión de Windows. Se muestran los 3 paneles principales, los cuales se muestran amigables para cualquier usuario.

Aunque el funcionamiento de los códigos programáticos en R es el mismo que en RSudio, en este trabajo se prefiere el último, por su mayor facilidad de uso, integración y estructuración.

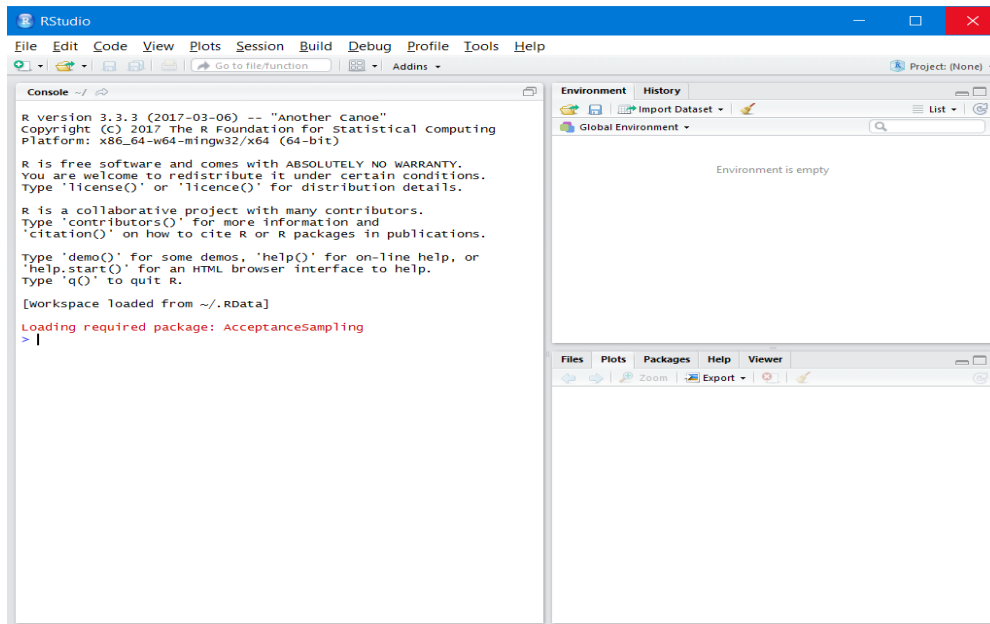


Figura 1.8 Pantalla inicial de Rstudio. Elaboración propia

2 Funcionamiento general de R y manejo de funciones

2.1 Uso de R

Este trabajo está orientado para lectores que ya tengan cierta experiencia en el manejo de R, por lo cual se abordaran sólo algunos conceptos del funcionamiento de este en un breve repaso.

R es un lenguaje de programación para el análisis y manipulación de datos, por lo que la programación se realiza por medio de funciones orientadas a objetos, lo que significa que las variables, datos, funciones, resultados, etc., se guardan en la memoria activa de la computadora en forma de objetos con un nombre específico. Se puede modificar o manipular estos objetos con operadores aritméticos, lógicos, y comparativos. En la tabla 2.1 se muestran los tipos de operadores aritméticos y comparativos de los que se puede hacer uso.

Operadores	
Aritméticos	Comparativos
+ adición	< menor que
- substracción	> mayor que
* multiplicación	<= menor o igual que
/ división	>= mayor o igual que
^ potencia	== igual
% / % división de enteros	!= diferente de

Tabla 2.1 Operadores en R. Elaboración propia con base en Paradise, (2003)

2.1.1 Ayuda

En R se puede buscar ayuda en línea, sólo se debe ingresar el siguiente signo “?” seguido de la función que se está buscando y automáticamente se despliega información detallada de la misma. Como ejemplo, ejecutando la siguiente línea de código, se mostrará información de cómo utilizar la función *plot*, a través de una ventana con información específica de la función, así como ejemplos de cómo se debe utilizar.

```
?plot
```

También se puede utilizar el comando *help* el cual tiene la misma función de brindar ayuda, como se muestra en la siguiente línea de código.

```
help(plot)
```

2.2 Tipos de variables y su asignación

Un objeto en R puede ser un vector, una lista, una función, arreglo, matriz, etc., que pueden ser de tipo numérico, carácter, complejo o lógico. Este puede ser creado con el operador “asignar” que es representado con los siguientes símbolos y operadores “<-“, “->” o “=”. A continuación, se darán 3 ejemplos de la forma en la cual se puede asignar.

Ejemplo 1

```
x <- c (10.4,5.6,3.1,6.4,21.7)
```

Ejemplo 2

```
c (10.4,5.6,3.1,6.4,21.7)-> x
```

Ejemplo 3

```
x = c (10.4,5.6,3.1,6.4,21.7)
```

En los ejemplos mostrados anteriormente se puede ver como un objeto en este caso llamado x guarda los siguientes valores (10.4, 5.6, 3.1, 6.4, 21.7). Posteriormente se puede realizar operaciones de adición, sustracción, división, etc., con dicho objeto, que a su vez puede ser mezclado con otros objetos para poder realizar la programación de un código que busca obtener información a través de datos reales.

2.2.1 Como manejar distribuciones de probabilidad

En R, manejar distribuciones de probabilidad no es una tarea complicada debido a que se pueden calcular fácilmente incluyendo probabilidades acumuladas, también maneja funciones de densidad y probabilidad puntual, así como la generación de valores pseudo-aleatorios siguiendo diferentes funciones de distribución habituales (tanto discretas como continuas). En la tabla 2.2. se muestran los nombres de varias funciones junto con argumentos adicionales que se pueden encontrar en R, en este caso para determinar números aleatorios con diferente distribución de probabilidad.

Distribución/función	función
Gausse (normal)	<i>rnorm (n, mean=0, sd=1)</i>
exponencial	<i>rexp (n, rate=1)</i>
gamma	<i>rgamma (n, shape, scale=1)</i>
Poisson	<i>rpois (n, lambda)</i>
Weibull	<i>rweibull (n, shape, scale=1)</i>
Cauchy	<i>rcauchy (n, location=0, scale=1)</i>
beta	<i>rbeta (n, shape1, shape2)</i>

Student' (t)	<i>rt (n, df)</i>
Fisher-Snedecor (F)	<i>rf (n, df1, df2)</i>
Pearson (χ^2)	<i>rchisq (n, df)</i>
binomial	<i>rbinom (n, size, prob)</i>
geométrica	<i>rgeom (n, prob)</i>
Hypergeométrica	<i>rhyper (nn, m, n, k)</i>
logística	<i>rlogis (n, location=0, scale=1)</i>
lognormal	<i>rlnorm (n, meanlog=0, sdlog=1)</i>
binomial negativa	<i>rnbinom (n, size, prob)</i>
uniforme	<i>runif (n, min=0, max=1)</i>
Estadístico de Wilcoxon's	<i>rwilcox (nn, m, n), rsignrank (nn, n)</i>

Tabla 2.2 Tipos de distribuciones en R. Recuperado de Paradise, (2003)

Para cada una de las distribuciones mostradas en la tabla anterior, se puede cambiar el prefijo de la función para obtener distinta información de los datos; los prefijos son los siguientes:

- **d** Función de densidad o de probabilidad puntual
- **p** Distribución acumulada
- **q** Función cuantil o percentil
- **r** Generar variables pseudo-aleatorias

Este trabajo está enfocado al control de la calidad de algún producto o servicio por lo cual se hará más énfasis en la distribución normal que es la que se ocupa usualmente para el análisis de control de calidad. En la tabla 2.3 se observan los prefijos anteriores para diferentes aplicaciones de la distribución normal.

Prefijo	Función	Descripción
d	dnorm (x, mean = 0, sd = 1, log = F)	Función de densidad
p	pnorm (q, mean = 0, sd = 1, lower. tail = T, log. p = F)	Distribución acumulada
q	qnorm (p, mean = 0, sd = 1, lower. tail = T, log. p = F)	Función cuantil o percentil
r	rnorm (n, mean = 0, sd = 1)	Generar variables pseudo-aleatorias

Tabla 2.3 Descripción de los prefijos de la distribución normal en R. Elaboración propia con base en Paradise, (2003)

La descripción de los argumentos que se muestran en cada una de las funciones de la tabla 2.3, son:

- **x, q**: Variable o vector de cuantiles.
- **p**: Vector de probabilidades.

- **n**: Números de observaciones.
- **mean**: Vector de medias. Por defecto, su valor es 0.
- **sd**: Vector de desviación estándar. Por defecto, su valor es 1.
- **log, log.p**: Parámetro booleano, si es TRUE, las probabilidades p son devueltas como $\log(p)$.
- **lower.tail**: Parámetro booleano, si es TRUE (por defecto), las probabilidades son $P[X \leq x]$, de lo contrario, $P[X > x]$.

En el siguiente código se muestra cómo encontrar 10 valores pseudo-aleatorios que siguen una distribución normal.

```
x <- rnorm (10)
## [1] -0.07990359 -0.06531329 -1.17337858 1.38286311 -0.28497070
## [6] -1.02290933 -1.51348260 1.47496178 0.47754027 2.71815436
0.126913017 0.134432286 0.355951466 0.009920816
```

La función de **dnorm** se utiliza para encontrar los valores de densidad en el eje x, que por medio de una expresión en R llamada “*curve*” que sirve para graficar una curva correspondiente a una función, en este caso **dnorm** que es una función de densidad y se utilizará en un intervalo de -3 a 3, como se muestra en la siguiente línea de código. El resultado al momento de ejecutar dicha línea de código se muestra en la figura 2.1.

```
curve(dnorm(x), from = -3, to = 3)
```

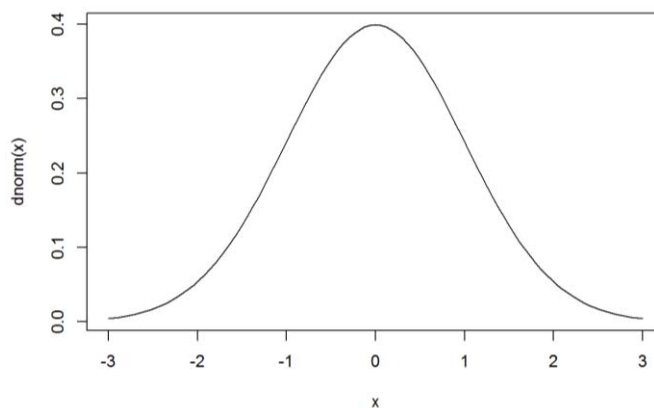


Figura 2.1. Curva de distribución normal. Elaboración propia

Para realizar una explicación del funcionamiento de las funciones **pnorm** y **qnorm** se tomará en cuenta que z es una variable aleatoria normal con una media de 0 y una desviación estándar igual a 1 (ejemplos 1 a 4).

Ejemplo 1. Con el código mostrado se realiza el cálculo de $P(Z > 2)$; se usa la función acumulada de distribución, indicando que la probabilidad de cola es hacia la derecha.

```
pnorm (2, mean = 0, sd = 1, lower.tail = FALSE)
## [1] 0.02275013
```

Ejemplo 2. Se muestra como al calcular $P(-2 \leq Z \leq 2)$ se vuelve a emplear la función de densidad acumulada, pero esta vez con la probabilidad de cola por defecto, hacia la izquierda.

```
pnorm(c(2), mean = 0, sd = 1) - pnorm(c(-2), mean = 0, sd = 1)
## [1] 0.9544997
```

Ejemplo 3. Para resolver $P(0 \leq Z \leq 1.73)$, se usan los mismos procedimientos anteriores.

```
pnorm(c(1.73), mean = 0, sd = 1) - pnorm(c(0), mean = 0, sd = 1)
## [1] 0.4581849
```

Ejemplo 4. Ahora se debe obtener el valor de a para que se cumpla la probabilidad de $P(Z \leq a) = 0.5793$ por lo cual se usa la función de cuantiles

```
qnorm(0.5793, mean = 0, sd = 1)
## [1] 0.20010
```

2.3 Leer datos desde un archivo

Aunque hay otras formas de lectura para diferentes formatos, para el uso de este trabajo, se utilizará *read.csv*, que sirve para leer archivos en excel, pero este archivo debe estar guardado con la extensión *.csv* para que puedan generar los datos en R.

A la instrucción *read.csv* se le puede dar el nombre del archivo para que lo busque en el ordenador sustituyendo la palabra *file* por el nombre exacto.

```
read.csv (file, header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, ...)
```

En caso de que no se conozca el nombre del archivo o su ubicación, se puede utilizar *file.choose*, que al momento de ejecutar el código despliega una ventana en la cual se busca el archivo con extensión *.csv* en la computadora, se selecciona y automáticamente

se cargan los datos en R, al usar *header* como se muestra en la instrucción anterior puede asignarle una variable lógica (FALSE o TRUE), indicando si el archivo contiene el nombre de las variables en la primera fila o línea.

Al ejecutar la instrucción:

```
línea1 <- read.csv(file.choose(), header = TRUE)
```

aparece un recuadro de dialogo para seleccionar el archivo.

2.3.1 Graficar

La manera en la cual se presentan los datos estadísticos después de un análisis es importante para tener una buena comunicación de los resultados correspondientes dentro de una organización. Los gráficos sirven para comprender mejor el comportamiento de los datos; en el control de la calidad visualizar los datos en forma gráfica es importante para detectar variabilidad, patrones y tendencia de los datos en el proceso, por lo cual es de utilidad saber cómo realizar los gráficos con ayuda de R.

R es un software con una gran variedad de gráficos para proporcionar al usuario una herramienta para la presentación de los datos; también cuenta con una gran flexibilidad para realizar los mismos, desde gráficos muy simples hasta gráficos personalizados que pueden ser modificados en cualquier momento.

En este trabajo no es posible detallar todos los tipos de gráficos que se pueden realizar en R, debido a que son muy extensos y al mismo tiempo cada uno de ellos cuenta con un número de parámetros que se pueden personalizar dependiendo de qué tan detallado se requiera el gráfico, por lo cual en este capítulo solo se dará un resumen de los gráficos que se pueden generar y cuál es su función en R así como sus parámetros que se pueden modificar; en el siguiente capítulo generaran gráficos enfocados al control de la calidad y se realizarán desde el más básico hasta uno más complejo.

La tabla 2.4 muestran solo algunos tipos de gráficos que se pueden realizar en R que para fines de este trabajo solo se muestra 7, para información más detallada de cómo utilizar otros tipos de gráficos se puede consultar el documento de (Paradis, 2003).

plot (x)	graficar los valores de x (en el eje y) ordenado en el eje x
plot (x, y)	Gráfico bivariado de x (en el eje x) y y (en el y)
piechart (x)	gráfico circular tipo 'pie'
boxplot (x)	gráfico tipo caja-y-bigote'
hist (x)	histograma de las frecuencias de x

barplot (x)	gráfica de barras para los valores de x
qqnorm (x)	cuartiles de x con respecto a lo esperado bajo una distribución normal

Tabla 2.4 Funciones graficas en R. Elaboración propia con base en Paradise, (2003)

Los gráficos se pueden clasificar en gráficos de alto nivel y gráficos de bajo nivel. Los gráficos de alto nivel son funciones que generan un gráfico principal mientras que los gráficos de bajo nivel son funciones que afectan al gráfico principal ya existente. En la tabla 2.5 se muestran comandos que se clasifican como de bajo nivel.

<i>points (x, y)</i>	agrega puntos (se puede usar la opción <i>type =</i>)
<i>lines (x, y)</i>	igual a la anterior, pero con líneas
<i>text (x, y, labels, ...)</i>	agrega texto dado por labels en las coordenadas (x, y); un uso típico: <i>plot (x, y, type = "n"); text (x, y, names)</i>
<i>mtext (text, side = 3, line = 0, ...)</i>	agrega texto dado por <i>text</i> en el margen especificado por <i>side</i> (ver <i>axis ()</i> más abajo); <i>line</i> especifica la línea del área de graficado
<i>segments (x₀, y₀, x₁, y₁)</i>	dibuja una línea desde el punto (x ₀ , y ₀) hasta el punto (x ₁ , y ₁)
<i>arrows (x₀, y₀, x₁, y₁, angle=30, code=2)</i>	igual al anterior, pero con flechas desde (x ₀ , y ₀) si <i>code=2</i> , al punto (x ₁ , y ₁) si <i>code=1</i> , o en ambos si <i>code=3</i> ; <i>angle</i> controla el ángulo desde la base de la flecha hasta la punta de la misma
<i>abline (a, b)</i>	dibuja una línea con pendiente b e intercepto a
<i>abline (h = y)</i>	dibuja una línea horizontal en la ordenada y
<i>abline (v = x)</i>	dibuja una línea vertical en la abcisa x
<i>abline (lm.obj)</i>	dibuja la línea de regresión dada por <i>lm.obj</i> (ver sección 5)
<i>rect ((x₁, y₁, x₂, y₂)</i>	dibuja un rectángulo donde las esquinas izquierdas, derecha, superior e inferior están dadas por (x ₁ , x ₂ , y ₁ , y ₂) respectivamente
<i>polygon (x, y)</i>	dibuja un polígono uniendo los puntos dados por x y y
<i>legend (x, y, legend)</i>	agrega la leyenda en el punto (x, y) con símbolos dados por legend
<i>title ()</i>	agrega un título y opcionalmente un subtítulo
<i>axis (side, vect)</i>	agrega un eje en la parte inferior, izquierda, superior o derecha (<i>side = 1, 2, 3, 4</i>); <i>vect</i> (opcional) da la abcisa u ordenada donde se deben dibujar los marcadores ('tick marks') del eje
<i>rug (x)</i>	dibuja los datos X en el eje x como pequeñas líneas verticales

Figura 2.5 Comandos de graficación de bajo nivel. Recuperado de Paradise, (2003)

2.3.2 Paquetes y complementos

Los paquetes en R son un conjunto de códigos en un formato estandarizado sobre algún tema en específico que sirven para ejecutar ciertas tareas de una manera más fácil y rápida. Estos paquetes son elaborados por la comunidad de usuarios y actualmente existen miles de paquetes que fueron realizados por usuarios de varias partes del mundo. Los paquetes se almacenan en un lugar llamado biblioteca dentro de R; para poder usar un paquete en R se debe tener conexión a internet.

R viene con un conjunto estándar de paquetes. Otros están disponibles para su descarga e instalación. Una vez instalado dicho paquete tiene que ser cargado en la sesión a utilizar.

En la figura 2.2 se muestra la pantalla de Rstudio en la cual se observa la pestaña con el nombre de “Packages” la cual contiene paquetes predeterminados con los que cuenta R sin necesidad de que estos sean descargados e instalados.

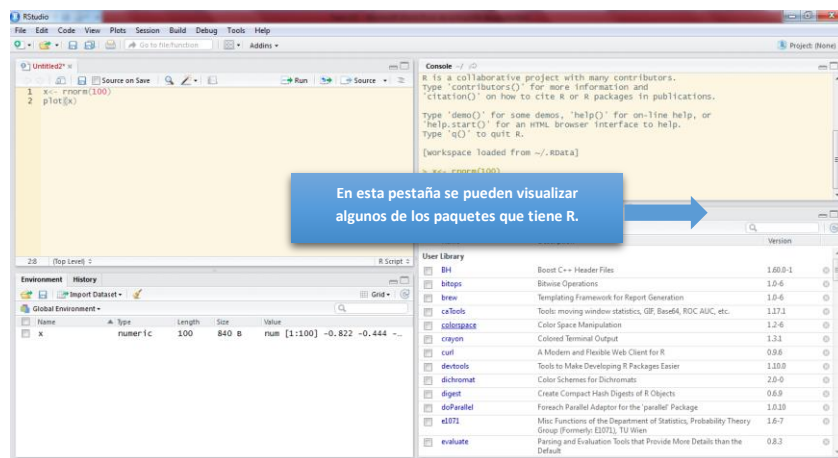


Figura 2.2. Pantalla de R Studio para instalar Packages. Elaboración propia

A continuación, se mostrará como instalar paquetes en Rstudio, los cuales son de utilidad para el usuario. El paquete por instalar lleva el nombre de *SixSigma*, el cual sirve para llevar a cabo análisis estadísticos en la forma de Seis Sigma. A través del ciclo DMAIC (Definir, Medir, Analizar, Mejorar y Controlar), también puede gestionar varios estudios de gestión de calidad tales como R & R, Análisis de capacidad, gráficos de control entre otros.

Pasos para instalar un paquete.

1. El usuario debe contar con el nombre del paquete a instalar, en este caso se instalará un uno llamado “SixSigma”. Se debe tener cuidado en respetar las letras mayúsculas y minúsculas que aparecen en el nombre de este, en caso contrario no se encontrará y por consecuencia no se instalará. Se debe ingresar la siguiente línea de código; es necesario que el nombre vaya entre comillas.

```
install.packages("SixSigma")
```

2. Se debe ejecutar la línea de código y empezará a descargarse el paquete como se muestra en la figura 2.3.

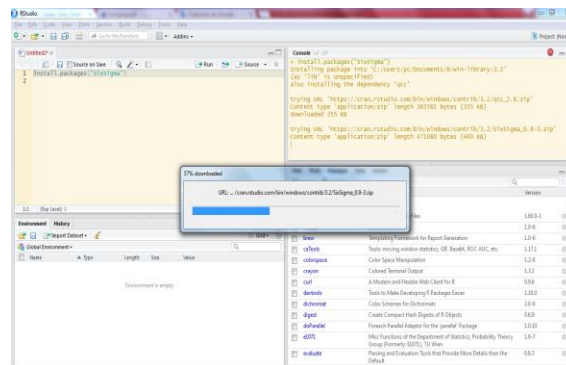


Figura 2.3. Instalación de paquete en R. Elaboración propia

3. Para poder utilizar las funciones y datos de la librería, se carga en el espacio de trabajo de R; esto se hace con la siguiente línea de código.

```
library("SixSigma")
```

4. Para verificar que ya se encuentra descargado e instalado en el espacio de trabajo como se muestra en la figura 2.4 al momento de colocar el nombre en la pestaña de búsqueda se puede observar que aparece nombre del paquete y una palomita lo cual significa que ya está listo para usarse.

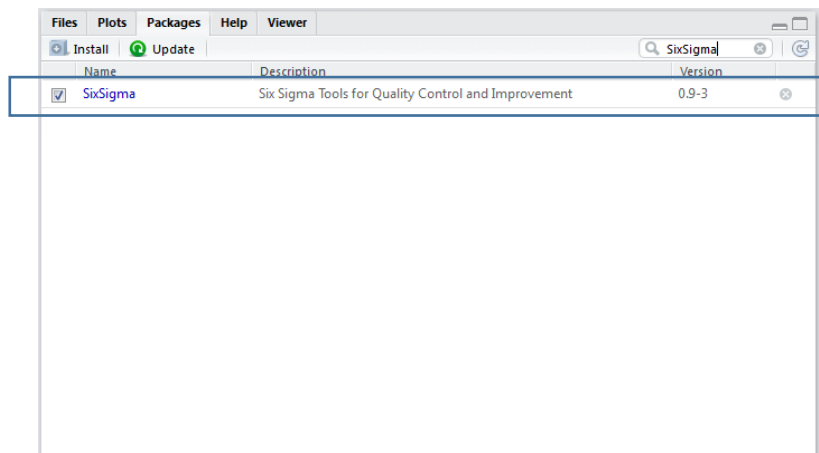


Figura 2.4. Espacio de trabajo con el paquete instalado. Elaboración propia

5. Por último, en dado caso que se requiera ser desinstalado, se utiliza la siguiente línea de código.

```
update.packages("SixSigma")
```

3 Generación de los códigos aplicables

3.1. Descripción del proceso hipotético analizado

Como ya se mencionó en el marco teórico, al planear los aspectos de calidad en la manufactura y/o en servicios, es importante asegurarse que el proceso cumple con las especificaciones y técnicas deseadas. Para iniciar el análisis se utilizará un caso de estudio como referencia, el cual se describe a continuación.

En este capítulo se muestra la estructura de un código hecho en R para el análisis estadístico de un proceso de producción, cabe mencionar que hay muchas formas de realizar el análisis en R, todo depende de la práctica y conocimiento del usuario sobre el software. Esta es una propuesta de las muchas que existen o pueden ser diseñadas para un usuario no muy experto.

El código se desarrolla con una empresa hipotética que lleva por nombre Fideos & Pastas, SA de CV, que quisiera certificar su proceso de producción de pasta. La empresa estableció que los paquetes de pasta de 200 g ya sean letras, balas, plumas o lo que sea, deben contener entre 192 y 208 gramos del producto. La empresa cuenta con 3 líneas de producción; cada línea produce una forma diferente de pasta.

Para analizar el comportamiento del proceso de empaque, el ingeniero de calidad Jorge Martinez, de la empresa Fideos & Pastas, decidió obtener el peso neto de los paquetes empacadas mediante muestreos horarios de 5 muestras cada hora, durante 24 horas en cada una de las 3 líneas de producción. La información de los muestreos de cada línea se muestra a continuación en la tabla 3.1.

Linea 1					Linea 2					Linea 3				
X1	X2	X3	X4	X5	X1	X2	X3	X4	X5	X1	X2	X3	X4	X5
194	195	198	196	205	202	193	189	185	189	188	185	208	185	193
197	193	200	205	191	206	189	190	189	202	189	186	200	185	195
204	205	203	209	200	199	209	195	190	206	214	188	188	185	212
201	195	195	190	205	198	185	197	190	205	205	189	204	188	213
188	189	194	193	187	197	185	202	191	195	190	190	175	189	184
190	202	193	207	201	198	187	197	192	213	185	190	197	190	186
192	206	188	185	210	197	187	195	193	211	193	195	190	195	187
195	191	196	194	195	188	190	204	193	193	188	196	192	197	208
168	194	195	210	214	194	190	202	193	188	213	197	206	202	212
200	204	202	204	208	213	191	195	195	213	185	197	199	195	207
189	206	204	208	204	199	191	185	196	197	188	198	202	195	211
185	206	195	211	208	189	192	197	196	189	211	199	191	195	214
201	207	204	207	198	185	193	185	200	203	208	199	210	197	205
204	209	199	205	195	191	194	185	203	199	189	205	199	202	200
189	211	192	189	207	197	194	188	205	209	213	185	188	197	194
207	213	210	211	208	192	204	202	205	207	193	187	209	202	195
207	205	204	194	210	212	205	212	205	198	201	210	200	204	193
199	203	204	202	210	197	207	215	207	204	195	191	208	204	197
212	197	207	196	210	199	207	204	209	191	195	192	189	205	196
204	188	187	210	211	211	208	205	211	206	190	187	213	208	195
201	198	202	203	211	212	210	208	211	205	206	191	190	211	207
205	189	194	197	212	214	210	211	212	194	203	190	198	212	197
202	199	202	214	213	206	211	212	213	195	207	194	185	212	198
207	213	191	206	213	205	205	195	206	214	204	205	204	215	193

Tabla 3.1 Muestreo de la línea 1,2 y 3. Elaboración propia

Con base en los datos mostrados en las tablas 3.1, se realizará un análisis del proceso con códigos de programación diseñados y elaborados en R.

Los datos muestrales se obtuvieron a través de un proceso de muestreo aleatorio, es decir, no suelen ser constantes, por lo que es necesario que se indique el valor alrededor del cual se agrupan los datos; asimismo se puede determinar una medida que haga referencia a la variabilidad. En este sentido pueden examinarse varias características, siendo las más comunes: la tendencia central de los datos, la dispersión o variación con respecto a este centro, los datos que ocupan ciertas posiciones, la simetría de los datos y la forma en la que los datos se agrupan. Se describirá el análisis para la línea 1; para las líneas 2 y 3 se presentarán los resultados en el capítulo 4.

Como se menciona en el punto 2.1.4. respecto a la lectura de datos en R, primero se tienen que tener los datos guardado en un objeto, para eso primero se deben leer los datos desde un archivo en excel con extensión *csv* para que sean cargados a R y así realizar el análisis correspondiente.

El objeto que tendrá guardados los datos sobre los pesos de los empaques de la línea 1 llevará el nombre de *línea1*, de acuerdo con los datos mostrados en la tabla 3.1.

La función *attach* ayuda a vincular la información de cada columna con el nombre de la misma. Por ejemplo, en la siguiente línea de código si no se usara la función *attach* y se necesitara información de alguna columna se tiene que usar la función *línea1\$X1* para saber qué información contiene esa columna.

```
linea1 <-read.csv(file.choose())  
  
linea1$X1  
  
[1] 194 197 204 201 188 190 192 195 168 200 189 185 201 204 189 207  
207 199 212 204 201 205 202 207
```

Con la función *attach* el manejo de datos es más práctico ya que como se muestra en el siguiente código vincula todos los datos con su respectiva columna y además la función *names(línea1)* te dice el número de columnas existentes y el nombre de cada una de ellas. Así mismo la siguiente vez que se desee utilizar los datos de la columna *X1*, *X2*, *X3*, *X4*, *X5* solo bastará con colocar su nombre y la operación que se necesita realizar.

```
linea1 <-read.csv(file.choose())  
  
attach(linea1)  
  
names(linea1)  
  
[1] "X1" "X2" "X3" "X4" "X5"  
  
> X1*X2
```



```
[1] 37830 38021 41820 39195 35532 38380 39552 37245 32592 40800 38934
38110 41607 42636 39879 44091 42435 40397 41764

[20] 38352 39798 38745 40198 44091
```

Los datos se tienen en 5 columnas que representan las muestras que se tomaron cada hora, por lo cual se tiene que obtener un vector con el total de los datos llamado *vlinea1*.

```
vlinea1 <- c (X1, X2, X3, X4, X5)

 [1] 194 197 204 201 188 190 192 195 168 200 189 185 201 204 189 207
207 199 212 204 201 205 202 207 195 193 205 195 189

 [30] 202 206 191 194 204 206 206 207 209 211 213 205 203 197 188 198
189 199 213 198 200 203 195 194 193 188 196 195 202

 [59] 204 195 204 199 192 210 204 204 207 187 202 194 202 191 196 205
209 190 193 207 185 194 210 204 208 211 207 205 189

 [88] 211 194 202 196 210 203 197 214 206 205 191 200 205 187 201 210
195 214 208 204 208 198 195 207 208 210 210 210 211

[117] 211 212 213 213
```

Para saber cuántos datos tiene el vector *vlinea1* se obtiene de la siguiente forma:

```
length(vlinea1)

## [1] 120
```

El promedio aritmético se puede obtener como:

```
mean(vlinea1)

## [1] 200.8
```

La mediana muestral es el percentil de orden 0.50; esto quiere decir que tiene por debajo al menos la mitad de los datos y por encima la otra mitad. Se puede obtener de la siguiente manera:

```
median(vlinea1)

## [1] 202.0
```

El mínimo y el máximo se puede obtener como:

```
max(vlinea1)

## [1] 214.0

min(vlinea1)

## [1] 168.0
```

O bien

```
range(vlineal)
## [1] 168 214
```

El rango es el máximo valor menos el mínimo valor se obtiene de la siguiente manera:

```
max(vlineal)-min(vlineal)
## [1] 46
```

O bien

```
diff(range(vlineal))
## [1] 46
```

Una medida más robusta que el rango es el rango intercuartílico. Se define como la diferencia entre los percentiles de orden 0,75 y 0,25, es decir, el cuartil superior menos el cuartil inferior. Se puede obtener con:

```
IQR(vlineal)
## [1] 12
```

La función *summary* que es una función que proporciona una descripción básica de los datos. En concreto, da el mínimo, el primer cuartil, la media, la mediana, el tercer cuartil y el máximo.

```
summary(vlineal)
##      Min.   1st Qu.   Median     Mean   3rd Qu.   Max.
##  168.0   195.0    202.0    200.8   207.0    214.0
```

También se puede plantear cómo conseguir un percentil de orden p (con $0 < p < 1$) arbitrario. En este caso se utilizará $p = 0.27$ para saber el valor bajo el cual se encuentran el 27% de los pesos obtenidos.

```
quantile(vlineal,0.27)
##      27%
## 195.0
```

O bien $p = 0.80$ lo cual significa que el 80% de los pesos están por debajo de 208 gramos mientras que el 20% está fuera de especificación de acuerdo con el límite superior.

```
quantile(vlineal,0.80)      ##      80%      ## 208
```

Cuando $p = 0.25$ al percentil se le denomina cuartil inferior. Si $p = 0.75$ se denomina cuartil superior y también se pueden obtener de la siguiente forma:

```
quantile(vlinea1,probs = c(0.25,0.75))  
##      25%      75%  
##      195      207
```

Para medidas descriptivas de dispersión podemos obtener la varianza y la desviación estándar la cuales se pueden obtener de la siguiente manera:

```
var(vlinea1)  
## [1] 67.31513  
sd(vlinea1)  
## [1] 8.20458
```

Como se puede ver, manipular datos en R y realizar un análisis previo de medidas descriptivas se puede hacer de forma rápida y sencilla.

3.1.1. Generación de histogramas

Generar un histograma en R es una tarea muy simple, depende de que tan personalizado se requiera; este gráfico se realiza con la siguiente función:

```
hist(x, breaks = "Sturges",freq = NULL, include.lowest = TRUE, right =  
TRUE,  
      density = NULL, angle = 45, col = NULL, border = NULL, main =  
paste("Histogram of" , xname),xlim = range(breaks), ylim = NULL, xlab =  
xname, ylab,axes = TRUE, plot = TRUE, labels = FALSE, nclass = NULL,  
warn.unused = TRUE, ...)
```

Para más información sobre cómo usar los argumentos de la función se puede consultar (Correa & Nelfi, 2002).

Para poder generar un histograma se necesita tener un vector de valores; en este caso de utilizará el vector que tiene por nombre *vlinea1* el cual tiene 120 valores que son el peso de los paquetes de pasta. El código para generar un histograma es el que se muestra a continuación y se muestra en la figura 3.1.

```
hist(vlinea1)
```

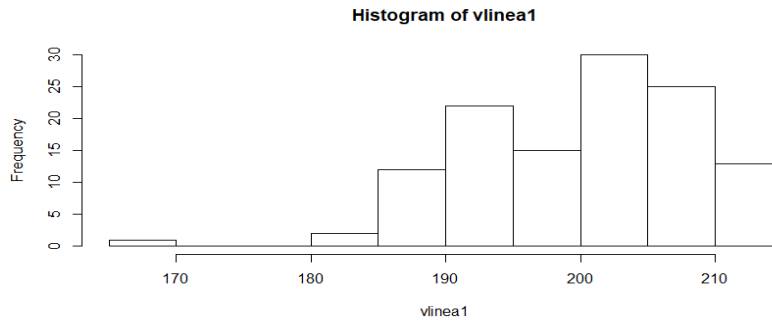


Figura 3.1 Histograma de peso de paquetes de pasta de la línea 1. Elaboración propia

Se puede personalizar el gráfico tanto como se desee, en la figura 3.2 se muestra un histograma con los mismos datos que el de la figura 3.1; a diferencia que el segundo se muestra con algunas modificaciones como; nombre de los ejes, color, título, y número de barras.

```
hist.default (vlinea1, breaks = 11, freq = TRUE,
             right = TRUE, col = "lightblue", border = "black",
             main = "Peso de paquetes de pasta",
             xlab = "Peso", ylab = "Frecuencia")
```

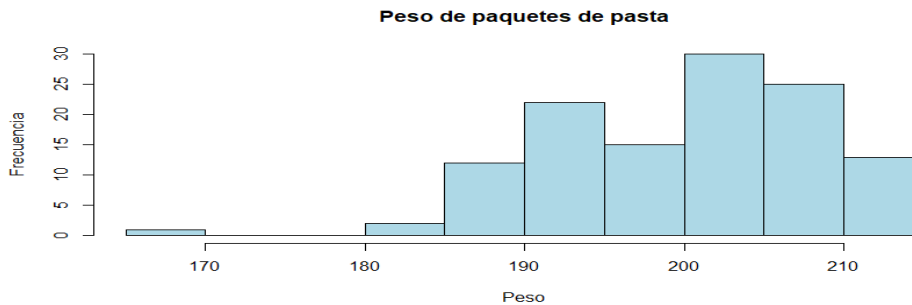


Figura 3.2 Histograma de peso de paquetes de pasta de la línea 1 personalizado. Elaboración propia

3.1.2. Prueba de hipótesis

En este sub capítulo, se realizará una prueba de hipótesis para validar la información que se tiene sobre el proceso de la línea 1, recordando que la empresa Fideos & Pastas, SA de CV, quiere certificar su proceso de producción por eso la empresa estableció que los paquetes de pasta deben pesar 200 gr., ya sean letras, balas, plumas o lo que sea. A la empresa no le conviene que los paquetes vayan con menos de 200 gr porque puede perder mercado, por otro lado, tampoco le conviene que lleven más de 200 gr por la pérdida económica que representa.

- Planteamiento de hipótesis:
 $H_0: \mu = 200$
 $H_a: \mu \neq 200$
- Nivel de significancia $\alpha = 0.05$

El código se debe iniciar con los datos generales como es el valor de alpha que representa el nivel de significancia, el valor hipotético (μ_0) y el valor de la media muestral (μ) que es un estimador de la media de la población, así como el tamaño de la muestra (n) y la desviación estándar ($sdTest$) de la muestra; estos valores se tienen que guardar en un objeto de la siguiente manera:

```
alpha <- 0.05
mu0 <- 200
mu <- mean(vlinea1)
n <- length(vlinea1)
sdTest <- sd(vlinea1)
```

Con base en la distribución muestral de un estadístico apropiado, se construye un criterio para probar la hipótesis nula contra la alternativa determinada. Se calcula, a partir de los datos, el valor estadístico sobre el cual se basa la decisión. Para trabajar con tablas normalizadas se usa Z que para función de este trabajo se llamará $ZTest$ que es el objeto que guarda el dato de 1.001372.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

También se calcula $Z_{\alpha/2}$ y $-Z_{\alpha/2}$ que forman parte de los límites de la región crítica o de rechazo, por medio de la función $qnorm$ que es una función que sirve para encontrar el valor de Z con una probabilidad de $\alpha/2$ que en este caso es igual a 0.025.

El argumento *lower.tail* significa si es *True* es $-Z_{\alpha/2}$ y si es *False* es $Z_{\alpha/2}$.

```
zTest <- (mu - mu0) / (sdTest/sqrt(n))
zalphaiz <- qnorm(alpha/2, lower.tail = True)
zalphader <- qnorm(alpha/2, lower.tail = False)
```

Para realizar el gráfico de la distribución normal se usa la función *pretty* que sirve para calcular una secuencia de números ($n+1$) a partir de los valores fijados, por ejemplo, en el eje X se necesitan valores de entre -3 y 3 con espacios de 0.1 que son 50 valores en total. Para el eje Y se busca una función de densidad en este caso *dnorm* con base en los valores de x .

Por último, para generar el gráfico de la distribución normal es por medio de la función *plot* con los valores de x y y , como se muestra en la figura 3.3.

```
x <- pretty(-3:3,n = 50)
y <- dnorm(x)
plot(x,y, ylab = "Densidad", type = "l", col = "black", main="Prueba de
hipotesis bilateral")
```

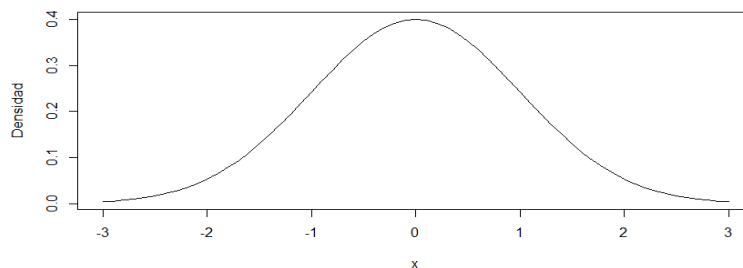


Figura 3.3 Gráfico de la distribución normal. Elaboración propia

Las siguientes líneas de código sirven para sombrear la región crítica en donde se rechaza la hipótesis nula, así como la región o zona de aceptación de la misma. Los valores que componen el objeto que lleva por nombre *cor.xi* son valores que van desde -3 hasta el valor de *zalphaiz* que es igual a -1.95996.

Dentro de este objeto se utiliza la función *seq* que sirve para generar números secuenciales dentro de un rango. Por ejemplo, en este caso se buscan números que vayan desde -3 hasta *zalphaiz* con una secuencia de 0.001, esto es con el fin de que cuando se generen los valores de densidad de *cor.yi* se pueda dibujar un polígono por medio de la función *polygon* cuyos vértices se dan en los valores de *cor.xi* y *cor.yi*. El gráfico sombreado con la primera región de rechazo de la hipótesis nula se muestra en la figura 3.4.

```
cor.xi <- c(-3,seq(-3,zalphaiz,by = 0.001),zalphaiz)
cor.yi <- c(0,dnorm(seq(-3,zalphaiz,by = 0.001)),0)
polygon(cor.xi,cor.yi,col = "cornflowerblue",lty = 1,lwd=1)
```

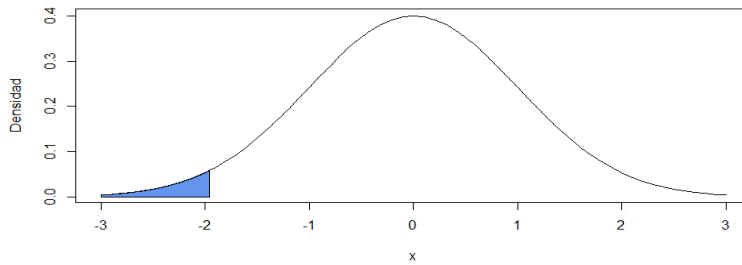


Figura 3.4 Gráfico con la zona de rechazo de la hipótesis nula. Elaboración propia

Para sombrear la zona de aceptación se utiliza el mismo código, solo se cambia el nombre de los objetos en este caso se llaman *cor.x* y *cor.y* que contienen los vértices para generar el polígono que va desde *zalphaiz* que es igual -1.95996 hasta *zalphader* que es igual a 1.95996 sobre el eje X. En la figura 3.5 se muestra el gráfico de acuerdo con las especificaciones en el código.

```
cor.x <- c(zalphaiz, seq(zalphaiz,zalphader, by=0.001),zalphader)
cor.y <- c(0,dnorm(seq(zalphaiz, zalphader, by=0.001)),0)
polygon(cor.x,cor.y, col = "cadetblue1",lty = 1,lwd=2)
```

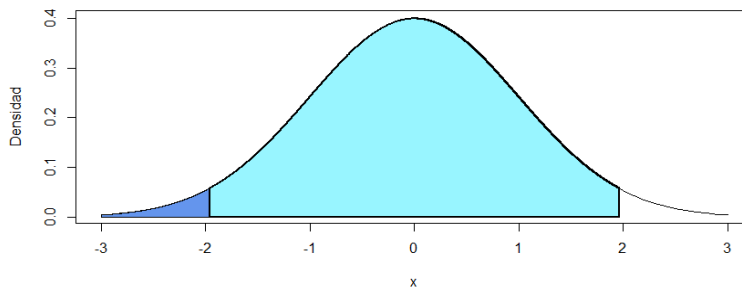


Figura 3.5 Gráfico con la zona de rechazo y aceptación de la hipótesis nula. Elaboración propia

Para concluir con el código se busca sombrear la segunda región crítica de rechazo de la hipótesis nula a través de la estructura de las mismas líneas de códigos anteriores a diferencia de que los valores sobre el eje X van desde *zalphader* que es igual 1.95996 hasta 3.

```
cor.xd <- c(zalphader,seq(zalphader,3, by=0.001),3)
cor.yd <- c(0,dnorm(seq(zalphader,3,by=0.001)),0)
polygon(cor.xd,cor.yd, angle = 45,col = "cornflowerblue",lty = 1,lwd=1)
```

La función *text* sirve para colocar texto al interior del gráfico, por medio de una ubicación sobre el eje X y Y, es importante mencionar que el texto a colocar debe ir entre comillas. Por último, la función *abline* sirve para colocar una línea al interior del gráfico desde una

posición sobre el eje X. En la figura 3.6 se muestra el gráfico completo con el código ejecutado en su totalidad.

```
text(zTest, 0.02, "ZTest", cex= NULL, pos=4, col="black")
abline(v = zTest, col = "black", lty=3)
text(0, 0.2, "Zona de aceptacion H0", cex= NULL, pos=1, col="black")
```

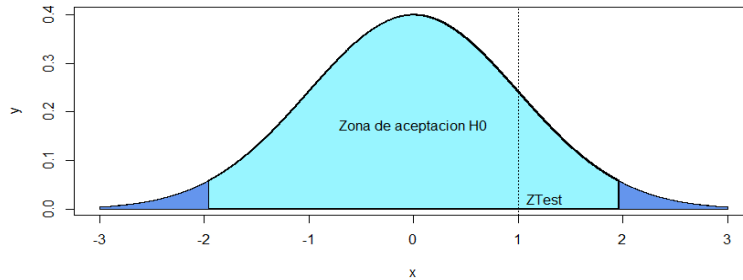


Figura 3.6 Gráfico con la zona de rechazo y aceptación de la hipótesis nula. Elaboración propia

Con ayuda del siguiente código se puede ver gráficamente si la hipótesis nula es rechazada o no es rechazada, en la figura 3.6 se puede ver que la hipótesis nula no es rechazada, lo cual refleja que la hipótesis H_0 está ampliamente respaldada por los datos de la muestra.

3.2. Herramientas básicas de la calidad

3.2.1. Diagrama de Pareto

El diagrama de Pareto se va a realizar con ayuda de un paquete de nombre “qcc” nombrado así por sus siglas en inglés “Quality Control Charts” el cual sirve para generar cartas de control de calidad Shewhart para datos continuos y atributos, cartas de control Cusum y EWMA. Índices de capacidad, gráfico de Pareto y gráfico de causa y efecto.

En el capítulo 2 en el apartado 2.1.5 se mostró como instalar un paquete por lo cual a continuación solo se describirá el código para realizar el mismo.

Los datos mostrados en la tabla 3.2 son los defectos de calidad encontrados en la empresa Fideos & Pastas SA de CV cuando se realizó una inspección del producto final.

Defectos	Calidad
Paquetes con producto de mas	80
Paquetes con producto de menos	59
Empaque roto	16
Cajas mal selladas	12
Cajas con faltante de producto	6

Tabla 3.2 Defectos de calidad. Elaboración propia

El archivo en Excel debe estar guardado con la extensión csv con los datos mostrados en la tabla 3.2.

Para este ejercicio el objeto lleva el nombre de “defectos” y se ejecuta el siguiente código:

```
defectos <- read.csv(file.choose(), header = TRUE)
attach(defectos)
names(defectos)
Variable <- Cantidad
names(Variable) <- Defecto
pareto.chart(Variable)
Tabla <- (pareto.chart(Variable))
```

Los datos obtenidos con el código se muestran a continuación en la figura 3.7 para poder construir el diagrama de Pareto mostrado en la figura 3.8.

Pareto chart analysis for variable

	Frequency	Cum.Freq.	Percentage	Cum.Percent.
Paquetes con producto de mas	80.000000	80.000000	46.242775	46.242775
Paquetes con producto de menos	59.000000	139.000000	34.104046	80.346821
Empaque roto	16.000000	155.000000	9.248555	89.595376
Cajas mal selladas	12.000000	167.000000	6.936416	96.531792
Cajas con faltante de paquetes	6.000000	173.000000	3.468208	100.000000

Figura 3.7 Datos del diagrama de Pareto

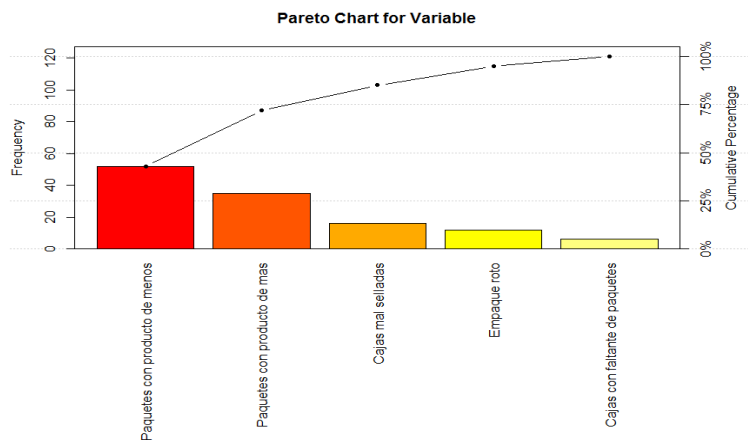


Figura 3.8 Diagrama de Pareto. Elaboración propia

3.2.2. Diagramas de dispersión

Un diagrama de dispersión está limitado a la presentación de dos variables. Para el análisis con los datos que se tienen de las líneas de producción no es necesario realizar un diagrama de dispersión, pero como herramienta para un análisis de control de calidad es de utilidad por lo cual se realizará un ejemplo de cómo utilizarlo en R.

En R se obtiene este gráfico mediante la función *plot*:

`plot(x, y, ...)` para revisar los argumentos que contiene dicha función se puede consultar (Correa & Nelfi, 2002).

Los siguientes datos que se analizaran corresponden al número de defectos encontrados en paquetes de pastas. Los datos se encuentran en la tabla 3.3.

Paquetes	Defectos
20	8
22	9
24	11
26	10
28	13
30	14
32	13
34	14
36	16
38	16
40	18
42	18

Tabla 3.3 Número de defectos. Elaboración propia

Por medio del siguiente código se puede crear un diagrama de dispersión para validar si existe tendencia entre el número de paquetes auditados contra el número de defectos encontrados en la muestra. Como primer punto se carga el archivo que como ya mencionamos tiene que estar guardado con la extensión csv en un objeto llamado *dispersión*, se utiliza la función *attach* para utilizar los nombres de las columnas con sus respectivos datos y asignarlos a un objeto llamado *x* para paquetes y *y* para defectos

El código mostrado en la parte inferior muestra el diagrama de dispersión mostrado en la figura 3.9.

```
dispersion <- read.csv(file.choose(), header = TRUE)
attach(dispersion)
names(dispersion)
x <- Paquetes
y <- Defectos
plot(x, y, type = "b", main = "Diagrama de dispersión",
     xlab = "Paquetes", ylab = "Defectos")
```

Diagrama de dispersion

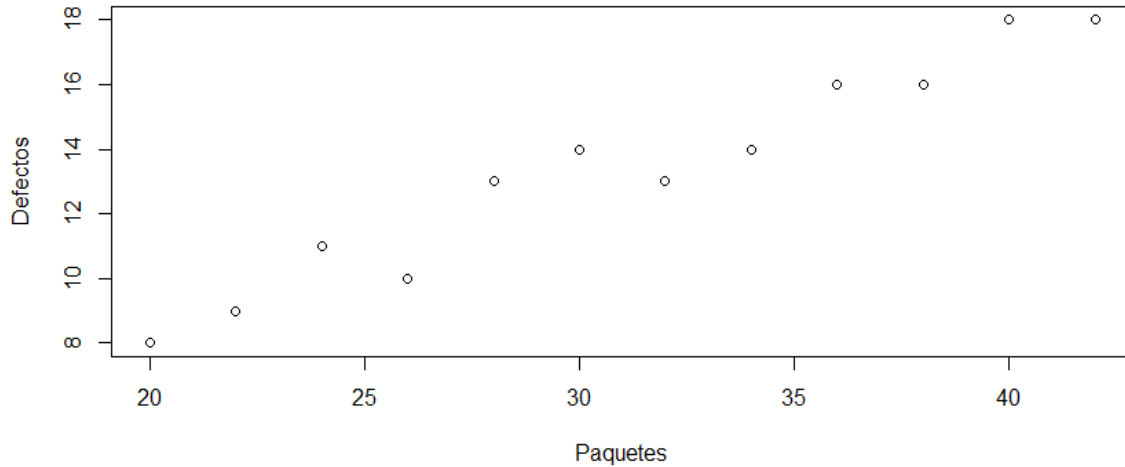


Figura 3.9 Diagrama de dispersión. Elaboración propia

3.3. Estudio de capacidad del proceso

A la empresa de Fideos & Pastas SA de CV, le interesa saber la capacidad de su proceso para poder determinar la variación que existe en el mismo de acuerdo con las especificaciones del cliente, en caso de no cumplir con las especificaciones realizar ajustes en el proceso para poder obtener la Certificación.

El análisis se iniciará trabajando con dos objetos que ya están cargados en R uno tiene por nombre *linea1*, este objeto tiene guardados los valores de los pesos de los paquetes muestreados en la línea 1 de forma matricial como se muestran en la tabla 3.1., y el otro lleva por nombre *vlinea1* que es un vector con el total de los datos de la línea 1.

Cabe mencionar que para iniciar este estudio de capacidad del proceso se tiene que validar que las variables de los pesos de los paquetes de pasta siguen una distribución normal. La función *qqnorm* es una función cuyo método predeterminado produce una gráfica Q-Q normal que significa un gráfico Cuantil-Cuantil que permite observar que tan cerca está la distribución de un conjunto de datos a alguna distribución ideal en este caso se quiere comparar con el gráfico de probabilidad normal. Por otro lado, la función *qqline* agrega una línea teórica sobre el grafico Q-Q normal que ayuda visualmente a saber si los datos siguen una distribución normal.

Por último, la función *Shapiro.test* se usa para contrastar si un conjunto de datos sigue una distribución normal o no, por medio de una prueba de hipótesis.

H_0 : los datos **proviene**n de una distribución normal

H_1 : los datos **no proviene**n de una distribución normal

Para determinar si los datos siguen una distribución normal se utiliza *p-value*, ya que si $p\text{-value} \leq 0.05$ se rechaza la hipótesis nula y si $p\text{-value} \geq 0.05$ no se rechaza la hipótesis nula.

Al ejecutar las siguientes líneas de código se mostrará un gráfico como el de la figura 3.10 así como también el valor de *p-value* que es menor a 0.05 ($0.0004056 < 0.05$) por lo tanto se rechaza la hipótesis nula H_0 , por lo que se puede afirmar que los datos no se distribuyen siguiendo una distribución normal.

```
qqnorm(vlineal)
qqline (vlineal, col = "red")
shapiro.test(vlineal)
```

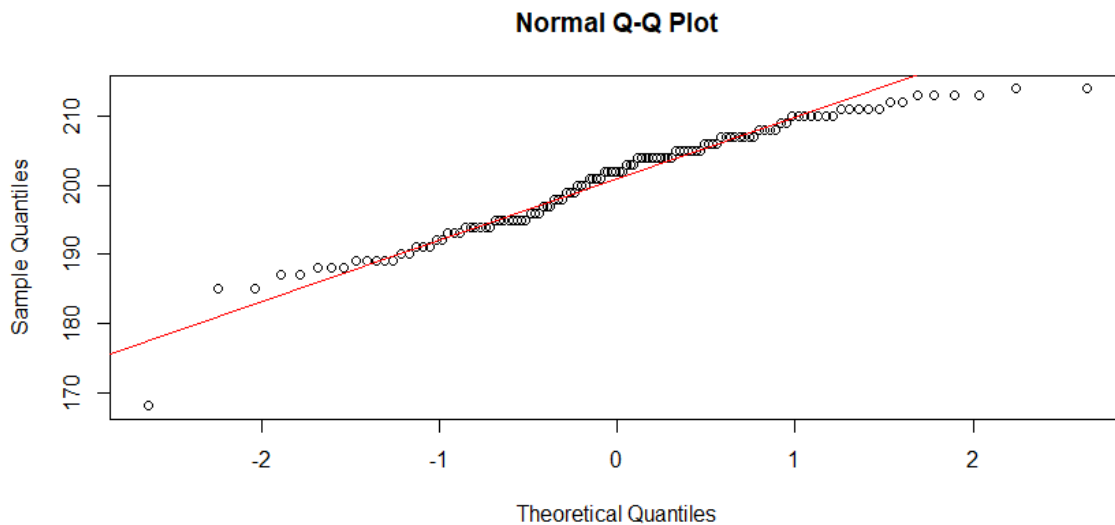


Figura 3.10 Gráfico de distribución normal. Elaboración propia

3.3.1 Estimación de μ_x y σ_x

Durante este análisis de estudio de la capacidad del proceso se utilizarán los valores estimados de μ_x y σ_x debido a que no se tienen datos históricos confiables sobre la población.

La estimación se puede realizar de diferentes maneras para estimar la media se utilizará el promedio de los 24 promedios de los 5 datos de cada hora muestreada. Así mismo, la desviación estándar, se estima por medio del promedio de los 24 valores de los rangos por cada hora muestreada, cabe mencionar que el promedio de los rangos (\bar{R}) es un estimador sesgado y se debe corregir por medio de una constante d_2 , esta constante varía dependiendo de n , en este caso $n=5$ por lo que $d_2 = 2.316$.

Para poder tener la estimación de la media se utilizará la función *apply* que sirve para trabajar datos en matrices y opera sobre renglones, columnas o ambos, por ejemplo, en las siguientes líneas de código se busca obtener el promedio, valor mínimo y máximo por cada renglón que representan los pesos de los paquetes muestreados por hora y en cada hora contiene 5 valores muestreados. La función *apply* trabaja sobre 3 argumentos de la siguiente manera:

```
apply(X, MARGIN, FUN, ...)
```

Donde:

X es un arreglo o matriz

MARGIN es una variable que sirve dependiente de cómo se quiere aplicar la función, MARGIN=1 (se aplica sobre renglones), MARGIN=2 (se aplica sobre columnas), MARGIN=c(1,2) (se aplica sobre columnas y renglones)

FUN es la función que quieres aplicar a los datos

De acuerdo con el código mostrado en la parte de abajo el objeto llamado *row.sums* contiene los promedios de cada hora, el objeto llamado *mini* guarda los valores mínimos de cada hora y el objeto *maxi* guarda los valores máximos de cada hora.

```
row.sums <- apply(linea1, 1, mean)
mini <- apply(linea1, 1, min)
maxi <- apply(linea1, 1, max)
```

Para poder realizar el cálculo de los índices de capacidad se necesita tener datos como son el límite inferior y superior de especificación, el valor nominal (N) del proceso que es el valor que se busca que cada paquete de pasta pese, así como también el valor estimado de la media que es el promedio de los promedios de cada hora del muestreo el cual esta guardado en un objeto llamado *mu*.

```
LES <- 208
LEI <- 192
mu <- mean(row.sums)
N <- 200
```

También se debe tener el estimador de la desviación estándar se utilizará el estimador

$$R_{pr}/d_2$$

Para el estimador de la desviación estándar se necesita calcular el rango de los valores de cada hora de la tabla de muestreo, así como el promedio de los rangos de cada hora y este valor dividirlo entre el factor d_2 .

```
rango <- maxi - mini  
d2 <- 2.316  
rdesvest <- mean(rango)/d2
```

Los índices de capacidad se calculan de acuerdo con las fórmulas que se explicaron en el punto 1.5 de este trabajo y las líneas de código que los calculan se muestran a continuación.

```
Cp <- (LES-LEI)/(6*rdesvest)  
Cr <- (6*rdesvest)/(LES-LEI)  
Cpi <- (mu-LEI)/(3*rdesvest)  
Cps <- (LES-mu)/(3*rdesvest)  
Cpk <- min(Cpi,Cps)  
#Índice de Taguchi  
tau <- sqrt((rdesvest^2)+((mu-N)^2))  
Cpm <- (LES-LEI)/(6*tau)
```

El nivel z es la métrica que sirve para calcular la distancia entre las especificaciones y la media del proceso. Cabe mencionar que entre mayor es z el proceso tiene menos productos fuera de especificación. Se utiliza z_i y z_s para encontrar la probabilidad de productos fuera de especificación por el lado izquierdo y el lado derecho respectivamente. Así como también encontrar el número de defectuosos por millón de oportunidad.

```
zi <- 3*Cpi  
Pzi <- 1-pnorm(zi)  
DPMOi <- Pzi*1000000  
zs <- 3*Cps  
Pzs <- 1-pnorm(zs)  
DPMOs <- Pzs*1000000
```

Por otro lado, el valor de z_{bench} es el resultado cuando todos los defectos se concentran en una cola de la distribución y como resultado se puede encontrar la probabilidad total de encontrar defectuosos en los paquetes de pastas.

```
zbench <- qnorm(1-(Pzi+Pzs))  
Pzbench <- 1-pnorm(zbench)  
DPMO <- Pzbench*1000000
```

Por último, para generar un resumen de todos los índices de capacidad se utiliza la función *data.frame* que sirve para almacenar tablas de datos y al momento de guardar esa tabla de datos en un objeto llamado Resumen se genera una tabla con todos los índices de capacidad.

```
Resumen<-  
data.frame(Cp,Cpi,Cps,Cpk,zi,Pzi,DPMOi,zs,Pzs,DPMOs,zbench,Pzbench,DPMO  
,Cpm)  
View(Resumen)
```

3.3.2 Representación gráfica de los resultados

Para representar gráficamente los resultados obtenidos en el punto 3.3.1. se iniciará graficando un histograma de los pesos de los paquetes de la línea 1 asignándolo a un objeto llamado *h* como se muestra en el siguiente código.

```
h <- hist(vlinea1, breaks = bins, freq = TRUE,  
          right = TRUE, col = "lightblue", border = "black",  
          main = "Peso de paquetes de pasta", xlim = c (170,230),  
          xlab = "Peso", ylab = "Frecuencia", las=1)
```

Se tienen que fijar los límites de especificación de los paquetes de pasta de 200 g que son 192 g y 208 g, por lo cual a través del siguiente código se fija dentro del histograma el límite inferior y superior de especificación, así como también el objetivo.

```
LES <- 208  
LEI <- 192  
N <- 200  
abline(v = LES, col = "red",lty=3)  
text(LEI, 30, "LEI", cex= NULL, pos=2, col="black")  
abline(v = LEI, col = "red",lty=3)
```

```

text(LES, 30, "LES", cex= NULL, pos=4, col="black")
abline(v = N , col = "red",lty=3)
text(N, 30, "Objetivo", cex= NULL, pos=2, col="black")

```

Por último, se dibuja la curva de distribución normal sobre el histograma en donde la función *lines* es una función que toma coordenadas dadas de varias maneras y une los puntos correspondientes con segmentos de línea. El objeto nombrado *normal* se multiplica por el multiplicador que es el resultado de la división de la frecuencia (*h\$counts*) entre la frecuencia relativa (*h\$density*)

```

myx <- seq(min(vlinea1), 230, length.out= 50)
mymean <- mean(vlinea1)
mysd <- sd(vlinea1)
normal <- dnorm(x = myx, mean = mymean, sd = mysd)
multiplier <- h$counts / h$density
lines (myx, normal * multiplier[1], col = "blue", lwd = 2)

```

La figura 3.11 muestra gráficamente los límites de especificación para la línea 1; la figura indica que una gran proporción de los paquetes se encuentran fuera de especificación.

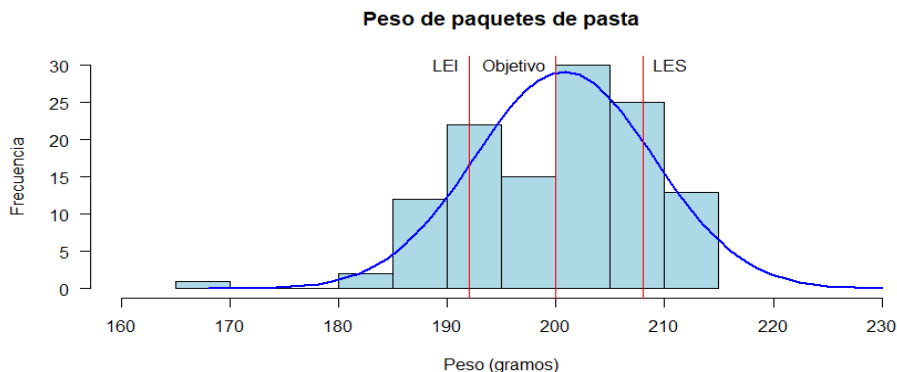


Figura 3.11 Límites de especificación línea 1. Elaboración propia

Con base en los resultados de z_i y z_s el 10.7% de los paquetes no cumple con la especificación inferior, mientras que el 15.2% no cumple con la especificación superior.

3.4. Diagramas de control

Para el desarrollo de este tema se utilizará la *figura 1.7 Diagrama de flujo para selección de gráficos de control*; como bien lo dice su nombre, servirá para seleccionar el gráfico de control a utilizar de acuerdo con el número de datos y tipo de datos.

Para iniciar con el análisis se debe verificar que se tienen cargados los datos de los pesos de los paquetes de pasta en el objeto llamado *linea1*. Como segundo paso se utilizará la

función *apply* que sirve para trabajar datos en matrices y opera sobre renglones, columnas o ambos como se mencionó en el punto 3.3.1. Se obtiene la media, la mediana, la varianza, el valor mínimo y el máximo de cada hora y se almacena la información en los objetos llamados *row.sums*, *mediana*, *varianza*, *mini* y *maxi*, respectivamente; así como también la desviación estándar y el rango.

```
row.sums <- apply(linea1, 1, mean)
mediana <- apply(linea1, 1, median)
varianza <- apply(linea1,1, var)
mini <- apply(linea1, 1 , min)
maxi <- apply(linea1, 1 , max)
desvest <- sqrt(varianza)
rango <- maxi - mini
```

Para continuar con el análisis se debe calcular el número de columnas y renglones que contiene la matriz de los datos de la línea 1 y eso se puede hacer por medio de las funciones *ncol* y *nrow*, como se muestra en el siguiente código. Cabe mencionar que el objeto *nc* es el tamaño de la muestra y *nr* es el número de veces que se muestreo.

```
nc <- ncol(linea1)
nr <- nrow(linea1)
```

Para estimar la media y la desviación estándar cabe mencionar que los dos son estimadores sesgados y se deben corregir por medio de una constante, esta constante varía dependiendo del tamaño de la muestra en este caso *nc*, por lo cual en un objeto llamado *factores* se guardara la información mostrada en la tabla 3.4 que son los valores de las constantes que se van a utilizar para realizar la estimación y poder construir los diagramas de control.

Tamaño de	<i>d2</i>	<i>d3</i>	<i>c4</i>
1			
2	1.128	0.853	0.7979
3	1.693	0.888	0.8862
4	2.059	0.88	0.9213
5	2.316	0.864	0.94
6	2.534	0.848	0.9515
7	2.704	0.833	0.9594
8	2.847	0.82	0.965
9	2.97	0.808	0.9693
10	3.078	0.797	0.9727
11	3.173	0.787	0.9754
12	3.258	0.778	0.9776
13	3.336	0.77	0.9794
14	3.407	0.763	0.981
15	3.472	0.756	0.9823
16	3.532	0.75	0.9835
17	3.588	0.744	0.9845
18	3.64	0.739	0.9854
19	3.689	0.734	0.9862
20	3.735	0.729	0.9869
21	3.778	0.724	0.9876
22	3.819	0.72	0.9882
23	3.858	0.716	0.9887
24	3.895	0.712	0.9892
25	3.931	0.708	0.9896

Tabla 3.4 Factores para la construcción de los diagramas de control. Elaboración propia

De acuerdo con el tamaño de *nc* que en este caso es 4 que es el número de muestras que se tomaron en cada hora, por medio del siguiente código se crean tres objetos con los nombres de *d2*, *d3* y *c4* en los cuales se guardan los valores de las constantes.

```
factores <- read.csv(file.choose(), header = TRUE)
d2 <- factores[nc,"d2"]
d3 <- factores[nc,"d3"]
c4 <- factores[nc,"c4"]
```

Tomando en cuenta la figura 1.7 de selección de diagramas de control y haciendo énfasis en los datos por variables, de acuerdo con el tamaño de la muestra que es igual a 4 se tiene que realizar un gráfico de control de tipo $X(\text{media})-R$. Para diseñar el gráfico de control se inicia con la estimación de la media y la desviación estándar.

```
muxr <- mean(row.sums)
sigmaxr <- (mean(rango))/(d2*sqrt(nc))
```

Los límites de control se calculan siguiendo las siguientes formulas:

$$\text{Limite inferior} = \bar{\bar{X}} - 3\left(\frac{\bar{R}/d_2}{\sqrt{n}}\right) \text{ y } \text{Limite superior} = \bar{\bar{X}} + 3\left(\frac{\bar{R}/d_2}{\sqrt{n}}\right)$$

```
LICxr <- muxr - (3*sigmaxr)
LSCxr <- muxr + (3*sigmaxr)
```

El siguiente código por medio del objeto llamado *limitesxr* se encuentran los valores mínimos y máximos de los límites de control y de los promedios de cada hora del muestreo, esto con el fin de fijar las escalas del eje y en el gráfico de control.

```
limLCxr <- c(min(LICxr),max(LSCxr))
limxr <- c(min(row.sums),max(row.sums))
limitesxr <- c(limLCxr,limxr)
```

Para realizar el gráfico de control se utiliza la función *plot*, tomando en cuenta algunos de los argumentos con los cuales cuenta dicha función.

```
plot(row.sums, ylim = c(min(limitesxr),max(limitesxr)), type = "b",
     col = ifelse((row.sums >= LSCxr | row.sums <= LICxr),
                  "red", "black"),
     pch = ifelse((row.sums >= LSCxr | row.sums <= LICxr), 17, 19), cex
     = 1,
     main = "Carta de control Xpr - R",
     xlab = "Subgrupo o muestra", ylab = "Promedio")
```

Para finalizar el código es importante colocar las líneas punteadas para identificar los límites de control, así como los nombres de cada línea como se muestra en la figura 3.12.

```
abline(h = LICxs, col = "green", lty = 3)
```

```

text(nr - 1, LICxs, paste('LIC =',round(LICxs,2)), col = "green", pos =
3)
abline(h = LSCxs, col = "blue", lty = 3)
text(nr - 1, LSCxs, paste('LSC =',round(LSCxs,2)), col = "blue", pos =
1)
abline(h = mean(row.sums), col = "red", lty = 3)
text(nr - 1, muxr, paste('mu =',round(muxr,2)), col = "red", pos = 3)

```

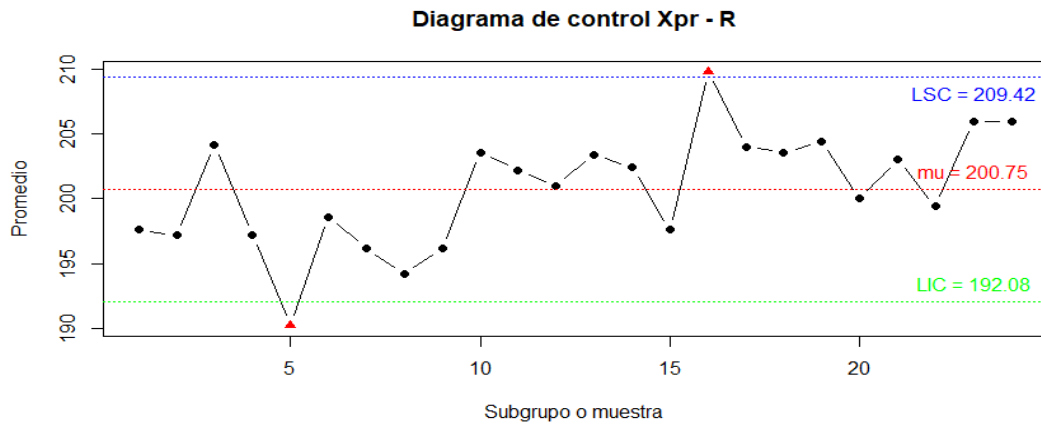


Figura 3.12 Diagrama de control X-R. Elaboración propia

3.4.1. Diagramas de control empleando la distribución binomial

Para realizar el análisis de estos diagramas de control se planeó determinar la cantidad de paquetes defectuosos (abiertos o mal cerrados) con muestras horarias de 60 cajas empacadas; sin embargo, el operador desafortunadamente tuvo distracciones no muestreó exactamente 60 paquetes cada hora (además pensaba que podía compensar su comportamiento muestreando más de 60 muestras de vez en cuando).

La información de los muestreos se muestra a continuación en la tabla 3.5.

Muestreo	n	Defectos	Defectos	Defectuosos
1	67	6	4	4
2	59	8	6	6
3	58	3	1	1
4	62	6	3	3
5	58	7	5	5
6	66	5	2	2
7	65	5	5	2
8	57	5	1	1
9	62	4	2	2
10	61	7	3	3
11	69	8	5	5
12	56	6	6	5
13	56	6	6	1
14	57	8	8	3
15	57	3	1	1
16	61	2	6	1
17	55	5	3	3
18	57	6	4	4
19	65	9	3	3
20	67	1	1	1
21	50	11	6	8
22	61	4	5	3
23	58	4	3	3
24	56	1	1	1

Tabla 3.5 Muestreo de defectos y unidades defectuosas. Elaboración propia

De acuerdo con el diagrama de flujo para selección de gráficos de control mostrado en la figura 1.7 y teniendo en cuenta que los tamaños de muestra no son constantes se aplicará la gráfica p (proporción o fracción de artículos defectuosos).

Como primer punto es guardar la tabla 3.7 en un objeto llamado *atributos1*, posteriormente se necesita convertir esa información en una matriz para poder separar por vectores cada columna como se muestra en el siguiente código.

```
atributos1 <- read.csv(file.choose(), header = TRUE)
atributos1 <- as.matrix(atributos1[1:5])
nr <- nrow(atributos1)
```

```
n<-as.vector(atributos1[,2])
Defectos.funcionales<-as.vector(atributos1[,3])
Defectos.en.acabado<-as.vector(atributos1[,4])
Defectuosos<-as.vector(atributos1[,5])
Defectos.totales <- Defectos.funcionales+Defectos.en.acabado
```

Para iniciar con la creación del gráfico de control p se debe calcular la proporción de defectuosos y se guarda en un objeto llamado p ; así como también la media y desviación estándar.

```
p <- Defectuosos/n
mup <- mean(p)
sigmap <- sqrt(p*(1-p)/n)
```

Siguiendo la misma estructura del código para la creación del diagrama de control X-R se inicia con el cálculo de los límites de control y el cálculo de los valores máximos y mínimos de los límites de control y de la proporción disconforme.

```
LICp <- mup - (3*sigmap)
LSCp <- mup + (3*sigmap)
limLCp <- c(min(LICp),max(LSCp))
limp <- c(min(p),max(p))
limitesp <- c(limLCp,limp)
```

Por medio de la función *plot* se crea el gráfico de control tomando en cuenta la escala del eje y para que el gráfico tenga una mejor visibilidad sobre los datos que muestra.

```
plot(p, ylim = c(min(limitesp),max(limitesp)), type = "b",
     col = ifelse((p >= LSCp | p <= LICp), "red","black"),
     pch = ifelse((p >= LSCp | p <= LICp), 17, 19), cex = 1,
     main = "Carta de control fracción disconforme (p)",
     xlab = "Subgrupo o muestra", ylab = "Fracción disconforme")
```

Por último, se agregan al gráfico los límites de control, así como el nombre de cada uno de ellos, se utiliza la función *point* debido a que los límites de control no son un valor fijo sino variable por los tamaños de muestra y el número de defectuosos encontrados en cada una de ellas.

```

points(LICp, type = "p", pch = 20, bg = par("bg"), col = "green", cex =
1)
text(nr - 1, min(LICp), paste("LIC"),col = "green")
points(LSCp, type = "p", pch = 20, bg = par("bg"), col = "blue", cex =
1)
text(nr - 1, max(LSCp), paste("LIC"),col = "blue")
abline(h = mean(p), col = "red", lty = 3)
text(nr - 1, mup, paste('mu =',round(mup,2)), col = "red", pos = 3)

```

En la figura 3.13 se muestra el diagrama de control

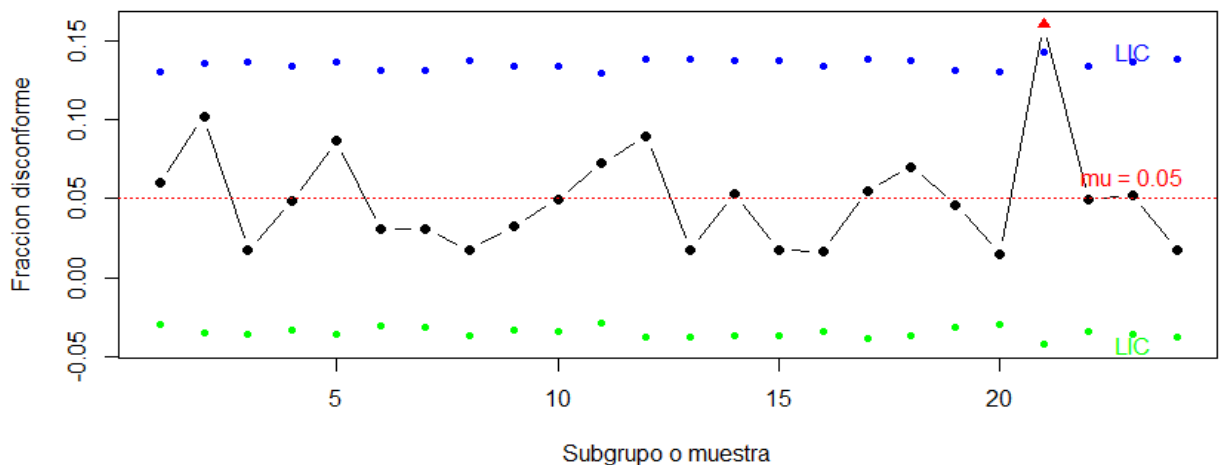


Figura 3.13 Diagrama de control fracción disconforme (p). Elaboración propia

Como parte del análisis es importante mencionar que los límites de control variables como se muestra en la figura 3.13 pueden dificultar la aplicación de los patrones especiales para la interpretación. Para evitar esa situación se puede usar un diagrama estandarizado en el cual, en lugar de graficar la proporción muestral, se grafica la proporción estandarizada. A continuación, el código para este diagrama de control.

Siguiendo la estructura que se ha venido ocupando para los diagramas de control, primero se tiene que calcular la media y desviación estándar, posteriormente los límites de control y por último el diagrama de control.

```

p <- Defectuosos/n
mup <- mean(p)
z <- ((p-mup)*sqrt(n))/(sqrt((1-mup)*mup))
LICz <- -3

```

```

LSCz <- +3
limLCz <- c(-3,3)
limz <- c(min(z),max(z))
limitesz <- c(limLCz,limz)
plot(z, ylim = c(min(limitesz),max(limitesz)), type = "b",
     col = ifelse((z >= LSCz | z <= LICz), "red","black"),
     pch = ifelse((z >= LSCz | z <= LICz), 17, 19), cex = 1,
     main = "Carta de control estandarizada para la fraccion
disconforme",
     xlab = "Subgrupo o muestra", ylab = "Zi")
abline(h = LICz, col = "green", lty = 3)
text(nr - 1, LICz, paste('LIC =',round(LICz,2)), col = "green", pos =
3)
abline(h = LSCz, col = "blue", lty = 3)
text(nr - 1, LSCz, paste('LSC =',round(LSCz,2)), col = "blue", pos = 1)
abline(h = 0, col = "red", lty = 3)

```

En la figura 3.14 se muestra el diagrama de control estandarizado

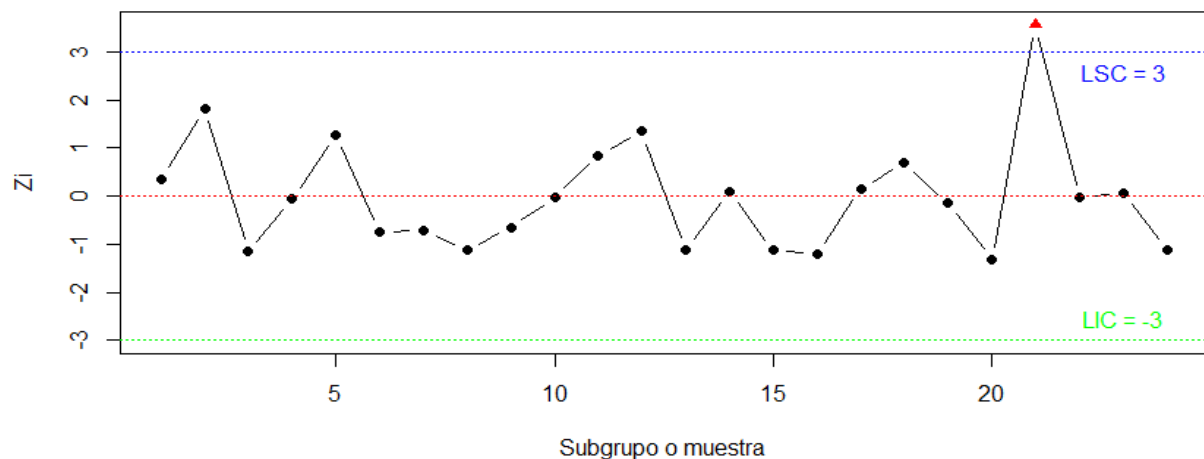


Figura 3.14 Diagrama de control estandarizado fracción disconforme. Elaboración propia

3.4.2. Diagramas de control empleando la distribución de Poisson

Para realizar el análisis de estos diagramas de control se planeó determinar la cantidad de defectos funcionales o de acabado encontrados en muestras horarias de 60 cajas empacadas; sin embargo, el operador desafortunadamente tuvo distracciones no muestreó exactamente 60 paquetes cada hora (además pensaba que podía compensar su comportamiento muestreando más de 60 muestras de vez en cuando).

Teniendo en cuenta que se va a trabajar con atributos, defectos no defectivos y que el tamaño de muestra no es constante, siguiendo el diagrama de flujo de la figura 1.7 se tiene que realizar un diagrama de control u (número de defectos por unidad). Para realizar el diagrama de control de defectos por unidad primero se inicia estimando la media y desviación estándar en donde n es el tamaño del subgrupo muestreado.

```
u <- Defectos.totales/n
muu <- sum(Defectos.totales)/sum(n)
sigmau <- sqrt(muu/n)
```

Como siguiente paso es calcular los límites de control, así como mínimos y máximo para definir la escala del eje y en el respectivo gráfico.

```
LICu <- muu - (3*sigmau)
LSCu <- muu + (3*sigmau)
limLCu <- c(min(LICu),max(LSCu))
limu <- c(min(u),max(u))
limitesu <- c(limLCu,limu)
```

Por último, se crea el gráfico de control con sus límites de control como se muestra en la figura 3.15.

```
plot(u, ylim = c(min(limitesu),max(limitesu)), type = "b",
     col = ifelse((u >= LSCu | u <= LICu), "red","black"),
     pch = ifelse((u >= LSCu | u <= LICu), 17, 19), cex = 1,
     main = "Diagrama de control para disconformidades por unidad (u)",
     xlab = "Subgrupo o muestra", ylab = "Disconformidades por unidad")
points(LICu, type = "p", pch = 20, bg = par("bg"), col = "green", cex =
1)
text(nr - 1, min(LICu), paste("LIC"),col = "green")
points(LSCu, type = "p", pch = 20, bg = par("bg"), col = "blue", cex =
1)
text(nr - 1, max(LSCu), paste("LIC"),col = "blue")
abline(h = mean(u), col = "red", lty = 3)
text(nr - 1, muu, paste('mu =',round(muu,2)), col = "red", pos = 3)
```

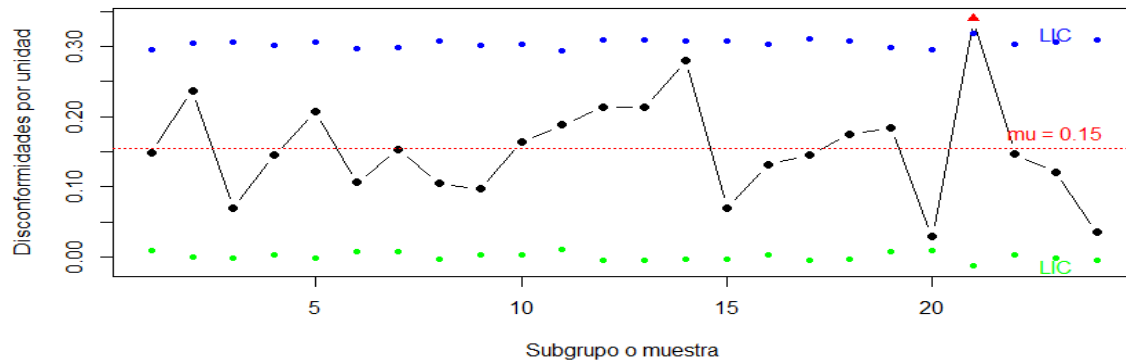


Figura 3.15 diagrama de control para disconformidades por unidad (u). Elaboración propia

En este caso los tamaños de los subgrupos no son muy distantes entre sí, pero en caso de encontrar una variabilidad más grande, se puede construir un diagrama estandarizado. La variable z sigue aproximadamente una distribución normal estándar ($\mu=0, \sigma=1$). Por lo tanto, la línea central en esta carta es igual a cero y los límites de control inferior y superior son -3 y 3 , respectivamente.

A continuación, el siguiente código sirve para generar el diagrama de control estandarizado mostrado en la figura 3.16 que se puede utilizar cuando los valores son graficados en unidades de desviación estándar y se debe calcular la z de prueba. Estos gráficos de control tienen su línea central en cero y sus límites de control inferior y superior son respectivamente -3 y 3 .

```

u <- Defecos.totales/n
muu <- sum(Defecos.totales)/sum(n)
z <- ((u-muu)*sqrt(n))/(sqrt((1-muu)*muu))
LICz <- -3
LSCz <- +3
limLCz <- c(-3,3)
limz <- c(min(z),max(z))
limitesz <- c(limLCz,limz)
plot(z, ylim = c(min(limitesz),max(limitesz)), type = "b",
     col = ifelse((z >= LSCz | z <= LICz), "red","black"),
     pch = ifelse((z >= LSCz | z <= LICz), 17, 19), cex = 1,
     main = "Diagrama de control estandarizada para las disconformidades
por unidad",
     xlab = "Subgrupo o muestra", ylab = "Puntuacion z")

```

```

abline(h = LICz, col = "green", lty = 3)

text(nr - 1, LICz, paste('LIC =',round(LICz,2)), col = "green", pos =
3)

abline(h = LSCz, col = "blue", lty = 3)

text(nr - 1, LSCz, paste('LSC =',round(LSCz,2)), col = "blue", pos = 1)

abline(h = 0, col = "red", lty = 3)

```

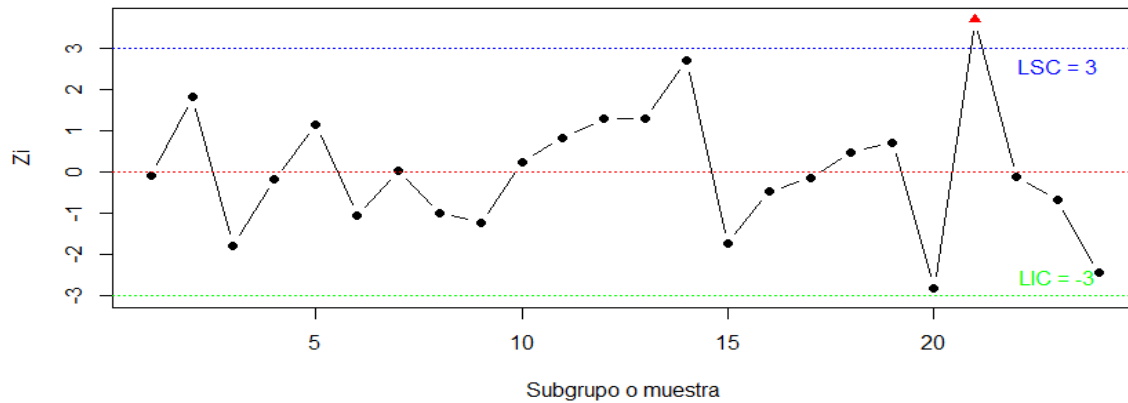


Figura 3.16 Diagrama de control estandarizado para las disconformidades por unidad. Elaboración propia

3.4.3. Diagramas CUSUM y EWMA

Los siguientes diagramas de control sirven para detectar cambios pequeños en el proceso, donde un cambio de nivel se considera pequeño si es menor a 1.5 veces el error estándar ($<1.5\sigma_{\bar{X}}$).

El diagrama de control CUSUM sirve para monitorear la media del proceso no se basa en las observaciones individuales, o promedios de una muestra de ellas, sino en la acumulación de las desviaciones de cada observación respecto a media global.

El siguiente código sirve para obtener el diagrama de CUSUM tabular que funciona acumulando las desviaciones de μ_0 que están arriba del objetivo con un estadístico C^+ y acumulando las desviaciones de μ_0 que están abajo del objetivo con otro estadístico C^- a estos estadísticos se les llama las cusums unilaterales superior e inferior, respectivamente. Se calculan de la siguiente forma:

$$C_i^+ = \text{máx}[0, x_i - (\mu_0 + K) + C_{i-1}^+]$$

$$C_i^- = \text{máx}[0, (\mu_0 - K) - x_i + C_{i-1}^-]$$

donde los valores iniciales son $C_0^+ = C_0^- = 0$

De acuerdo con (Montgomery, 2016) la cusum tabular se diseña eligiendo valores para el valor de referencia K y para el intervalo de decisión H , se han realizado diversos estudios analíticos del desempeño de la ARL (longitud promedio de corrida) para seleccionar mejor los valores de H y K .

Se define $H = h\sigma$ y $K = k\sigma$ donde σ es la desviación de la variable muestral usada para formar la cusum. Utilizando $h = 4$ o $h = 5$ y $k = 1/2$ por lo general se obtendrá una cusum que tiene propiedades convenientes de la ARL que corresponde al número promedio de puntos que deben ser graficados antes de que un punto indique una condición fuera de control, para un corrimiento de 1σ se detectaría en 8.38 muestras con ($k = 1/2$ y $h = 4$) o en 10.4 muestras con ($k = 1/2$ y $h = 5$) que en comparación con una carta de control de Shewhart para mediciones individuales requeriría 43.96 muestras, en promedio, para detectar este corrimiento.

Para este caso utilizaremos $h = 4.77$, debido a que se obtendrá una cusum con $ARL_0 = 370$ muestras el cual coincide con los límites 3σ usuales con una probabilidad de 0.0027 de que este fuera de control.

Por otro lado, en este caso se está trabajando con subgrupos racionales simplemente se sustituye x_i con \bar{x}_i (promedio muestral o de los subgrupos) en las fórmulas anteriores, y σ se sustituye con $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

En el siguiente código se inicia teniendo el valor objetivo μ_0 así como también el valor de σ que para este caso vamos a estimar $\sigma_{\bar{x}} = \frac{\bar{R}/d_2}{\sqrt{n}}$.

```
mu0 <- 200
sigmax <- mean(rango)/d2 * 1/sqrt(nc)
K <- sigmax*0.5
```

Para obtener los valores del estadístico C^+ y C^- se crea un objeto llamado Cs y Ci respectivamente, en el cual se van a ir guardando los valores que se generen con la función *for*; como se muestra en el código de abajo en el cual va guardando la suma acumulada de cada interacción entre el valor objetivo y la media de cada subgrupo.

```
Cs <- c()
Cs <- max(0, row.sums[1]-(N+K)+Cs[1])
for (i in 1:nr+1) {
  Cs [i] <- max(0, row.sums[i-1]-(N+K)+Cs [i-1])
}
```

```

Ci <- c()
Ci <- max(0,(N-K)-row.sums[1]+Ci[1])
for (i in 1:nr+1) {
  Ci[i] <- min(0,row.sums[i-1]-(N-K)+Ci[i-1])
}

```

Como en los diagramas de control anteriores se calculan los límites de control inferior y superior que en este caso están dados por el intervalo de decisión, $LICc = -H$ y $LSCc = H$.

```

LICc <- (-4.77)(sigmax)
LSCc <- (4.77)(sigmax)

```

Para fijar las escalas del eje en el diagrama de control se calculan los valores mínimos y máximos de los límites de control y de los estadísticos C^+ y C^- , y se guardan en un objeto llama *limitesc*.

```

limLCC <- c(min(LICc),max(LSCc))
limCs <- c(min(Cs),max(Cs))
limCi <- c(min(Ci),max(Ci))
limitesc <- c(limLCC,limCs,limCi)

```

Para realizar el gráfico de control se utiliza la función *plot*, tomando en cuenta algunos de los argumentos con los cuales cuenta dicha función. Primero se grafica el estadístico C^+ .

```

plot(Cs, type = "b", col = ifelse((Cs > LSCc | Cs <
LICc),"red","black"),
      pch = ifelse((Cs > LSCc | Cs < LICc), 17, 19), cex = 1, main =
"Cusum",
      xlab = "Subgrupo o muestra",ylab = "Suma acumulada",
      ylim = c(min(limitesc),max(limitesc)))

```

Posteriormente se grafica el estadístico C^- en la misma gráfica, pero por medio de la función *lines*.

```

lines(Ci, col = ifelse((Ci > LSCc | Ci < LICc),"red" ,"black"), type =
"b",
      pch = ifelse((Ci > LSCc | Ci < LICc), 17, 19), cex = 1)

```

Para concluir con el gráfico que se muestra en la figura 3.17 se agregan las líneas de los límites de control con su nombre y el valor de cada uno de ellos.

```

abline(h = LICc, col = "green", lty = 3)
text(nr - 1, LICc, paste('LIC = ', round(LICc, 2)), col = "green", pos = 3)
abline(h = LSCc, col = "blue", lty = 3)
text(nr - 1, LSCc, paste('LSC = ', round(LSCc, 2)), col = "blue", pos = 1)
points(Ci, type = "b", col = ifelse((Ci >= LSCc | Ci <= LICc), "red",
  "black"),
      pch = ifelse((Ci >= LSCc | Ci <= LICc), 17, 19), cex = 1)
abline(h = 0, col = "red", lty = 3)

```

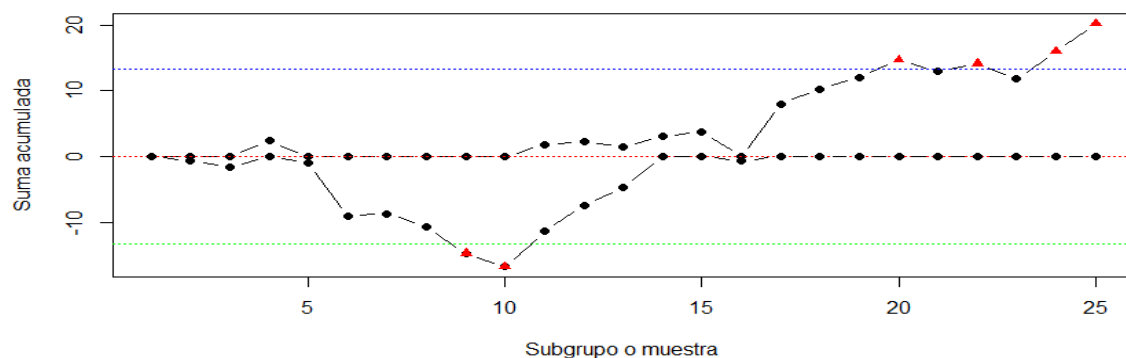


Figura 3.17 Diagrama de control CUSUM. Elaboración propia

Por otro lado, el diagrama de control EWMA se crea a través del promedio móvil ponderado exponencialmente y también sirve para monitorear la media del proceso.

Este tipo de diagrama se utiliza muy frecuentemente en el modelado de series de tiempo y al hacer pronósticos.

El código se iniciará guardando el valor de λ equivalente a 0.2 y muy similar a lo que hizo en el diagrama de cusum se crea un objeto con el nombre de Z_t y σZ_t para guardar los valores que se vayan generando a través de la función *for*. La fórmula es la siguiente:

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}$$

donde $Z_0 = \bar{X}$ que coincide con el valor nominal si el proceso está centrado, y $0 < \lambda \leq 1$. Por ejemplo, hasta Z_3 , las tres primeras sumas estarían dadas por:

$$Z_1 = \lambda \bar{x}_1 + (1 - \lambda)Z_0$$

$$Z_2 = \lambda \bar{x}_2 + (1 - \lambda)Z_1$$

$$Z_3 = \lambda \bar{x}_3 + (1 - \lambda)Z_2$$

Por lo tanto, el código se muestra de la siguiente manera:

```
lamda <- 0.2
Zt <- c()
sigmaZt <- c()
Zt[1] <- (lamda * row.sums[1]) + ((1 - lamda) * muxr)
lam <- sqrt((lamda)/(2 - lamda))
sigmaZt[1] <- sigmaxr * lam * sqrt((1 - (1 - lamda) ^ (2)))
for (i in 2:nr) {
  Zt[i] <- lamda*row.sums[i] + (1 - lamda) * Zt[i - 1]
  sigmaZt[i] <- sigmaxr*lam*(sqrt(1 - (1 - lamda) ^ (2*i)))
}
```

Se calculan los límites de control con base en las siguientes formulas:

$$LCS = Z_0 + 3 \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\left[\frac{\lambda}{2 - \lambda} \right] [1 - (1 - \lambda)^{2t}]}$$

$$LCI = Z_0 - 3 \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\left[\frac{\lambda}{2 - \lambda} \right] [1 - (1 - \lambda)^{2t}]}$$

Así como también se calculan los valores mínimos y máximos de los límites de control y el estadístico EWMA, que grafica al tiempo t en el diagrama, para fijar las escalas del eje y .

```
LIC <- muxr - (3*sigmaZt)
LSC <- muxr + (3*sigmaZt)
limLC <- c(min(LIC),max(LSC))
limZt <- c(min(Zt),max(Zt))
limites <- c(limLC,limZt)
```

Por medio de la función *plot* se realiza el grafico que se muestra en la figura 3.18.

```
plot(Zt, type = "b", col = ifelse((Zt > LSC | Zt < LIC), "red", "blue"),
     pch = ifelse((Zt > LSC | Zt < LIC), 17, 19), cex = 1, main =
"EWMA",
     xlab = "Subgrupo o muestra", ylab = "Zt",
     ylim = c(min(limites),max(limites)))
```

Los límites de control como se muestra en la figura 3.18 tienen una variación al inicio hasta que se estabilizan, por lo cual por medio de la función *points_se* puede dar ese efecto en el gráfico. Por último, se colocan los nombres de los límites de control y sus valores.

```
points(t,LIC,pch = "_", col = "red",lwd = 2)
points(t,LSC,pch = "_", col = "red")
abline(h = muxr, col = "red", lty = 3)
text(nr - 2, muxr, paste('mu =',round(muxr,2)), col = "black", pos = 3)
text(nr - 2, tail(LIC, n = 1), paste('LIC =',signif(tail(LIC, n =
1))),col = "black", pos = 3)
text(nr - 2, tail(LSC, n = 1), paste('LSC =',signif(tail(LSC, n =
1))),col = "black", pos = 1)
```

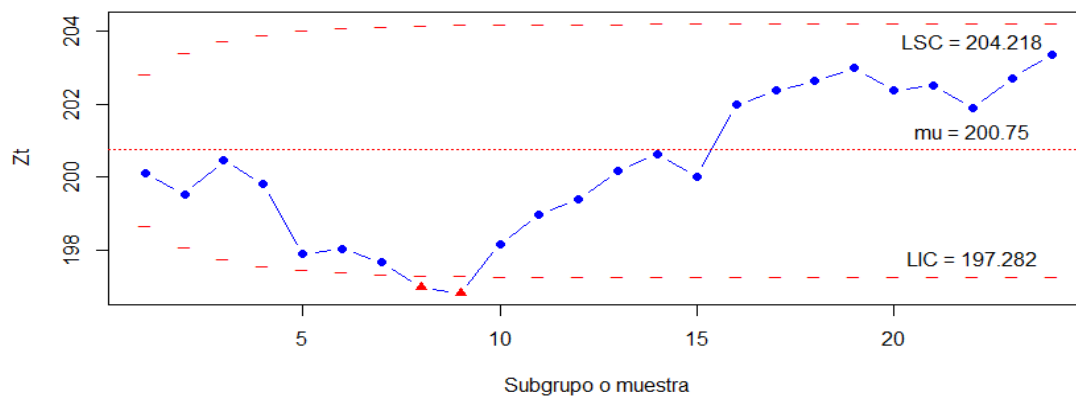


Figura 3.18 Diagrama de control EWMA. Elaboración propia

4 Análisis de control estadístico e interpretación de códigos generados

De acuerdo con los datos proporcionados de la línea 1,2 y 3 la estadística descriptiva se muestra en la tabla 4.1 como resumen de las 3 líneas.

Estadístico descriptivo	Línea 1	Línea 2	Línea 3
Media	200.75	199.36	197.55
Mediana	202	198	197
Máximo	214	215	215
Mínimo	168	185	175
Q1	195	192	190
Q3	207	206	205
IQR	12	14	15
Desviación estándar	8.02	8.63	8.98
Varianza	67.31	74.50	80.78

Tabla 4.1 Resumen de estadísticas generadas por Rstudio. Elaboración propia

De acuerdo con el resumen de la tabla 4.1 sobre los pesos de los paquetes de pasta se observa que, de las 3 líneas, la línea 3 presenta una mayor variabilidad entre los datos, y la línea 1 tiene una tendencia central muy aproximada al valor que busca la empresa que son 200 gramos por paquete.

Tomando en cuenta la desviación estándar se observa que la línea 3 es la de mayor dispersión en los datos de la muestra. En la figura 4.2 se muestran los histogramas de las

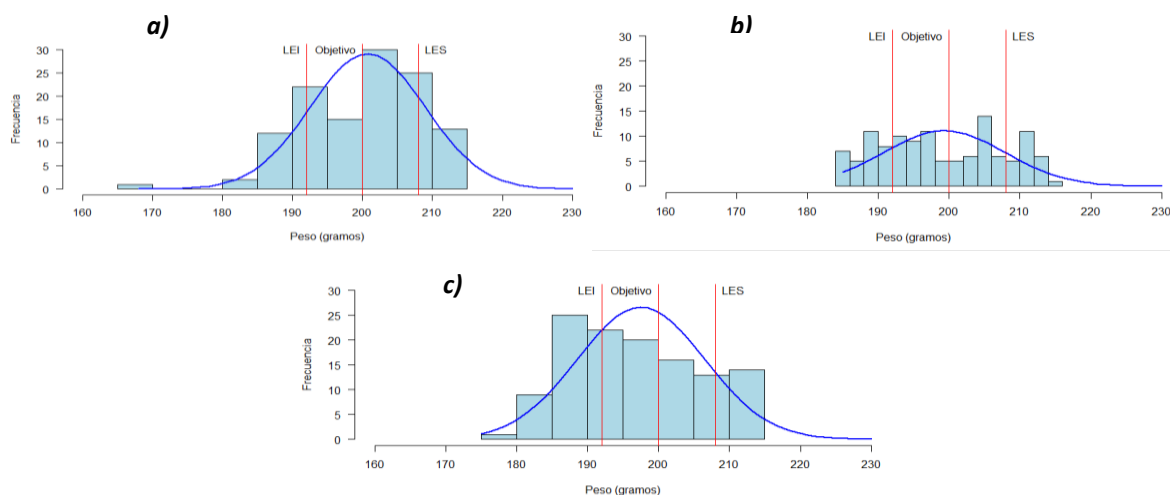


Figura 4.1 Histograma generado con Rstudio con límites de especificación y con 11 clases. a) línea 1, b) línea 2, c) línea 3. Elaboración propia

4.1 Prueba de hipótesis

Analizado la prueba de hipótesis para validar la información que se tiene sobre el proceso de las 3 líneas. A la empresa no le conviene que los paquetes vayan con menos de 200 gr porque puede perder mercado, por otro lado, tampoco le conviene que lleven más de 200 gr por la pérdida económica que representa.

- Planteamiento de hipótesis:
 $H_0: \mu = 200$
 $H_a: \mu \neq 200$
- Nivel de significancia $\alpha = 0.05$

Los valores del estadístico de prueba Z_0 para las 3 líneas se muestran en la tabla 4.2 así como también los valores $Z_{\alpha/2}$, y tomando en cuenta que se rechaza H_0 si $|Z_0| > Z_{\alpha/2}$, donde $Z_{\alpha/2}$ es el punto porcentual $\alpha/2$ superior de la distribución normal estándar.

Línea	$ Z_0 $	$Z_{\alpha/2}$	Conclusión
1	1.0013	1.9599	No existe evidencia para rechazar H_0
2	0.0837	1.9599	No existe evidencia para rechazar H_0
3	2.9859	1.9599	Sí existe evidencia para rechazar H_0

Tabla 4.2 Valores de prueba de hipótesis. Elaboración propia

En la figura 4.2 se observa gráficamente la zona de aceptación de H_0 así como también los intervalos $(Z_{\alpha/2}, \infty)$ y $(-\infty, -Z_{\alpha/2})$ que forman la región crítica de la prueba. Para la línea 2 como se muestra en la figura 4.2a no existe evidencia para rechazar la hipótesis nula, por lo tanto, se concluye que no hay evidencia suficiente que indique que el peso de los paquetes difiere de 200 gramos, en cambio para la figura 4.2b que corresponde a la línea 3 sí existe evidencia para rechazar la hipótesis nula, por lo cual se concluye que hay suficiente evidencia de que los paquetes de pasta diferentes de 200 gramos.

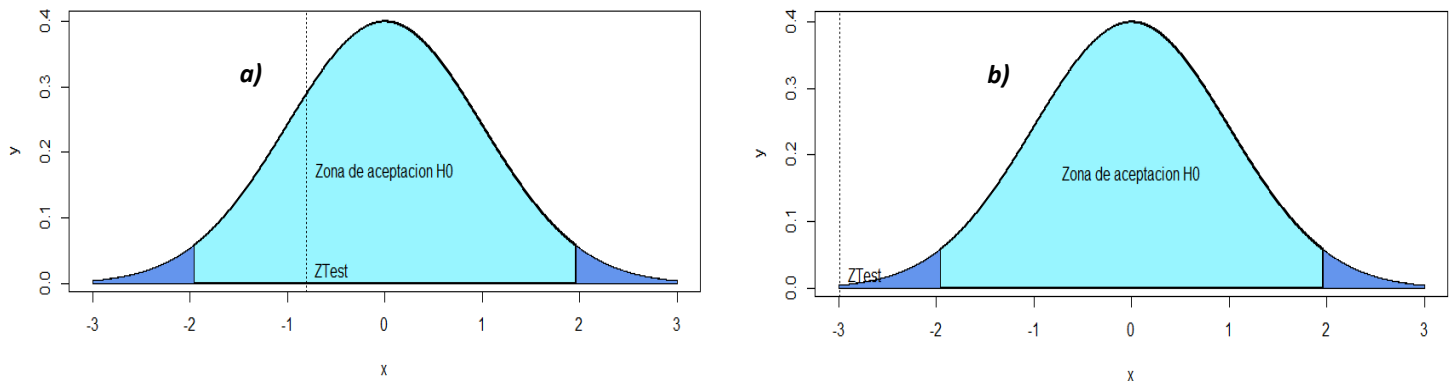


Figura 4.2 Distribución normal de prueba de hipótesis. a) línea 2, b) línea 3. Elaboración propia

4.2 Índices de capacidad

El análisis de capacidad de un proceso se realiza principalmente por dos objetivos, uno es verificar si el proceso es capaz de cumplir con las especificaciones del producto y el otro saber el por qué el proceso no cumple con las especificaciones del producto, entonces se habla que sirve como predicción y como análisis.

En la tabla 4.3 se muestra un resumen de los índices de capacidad de las 3 líneas de llenado de paquetes de pasta.

Índice de Capacidad	Línea 1	Línea 2	Línea 3
C_p	0.3353079	0.3245714	0.275682
C_{pi}	0.366743	0.2988762	0.1912544
C_{ps}	0.3038728	0.3502667	0.3601096
C_{pk}	0.3038728	0.2988762	0.1912544
Z_i	1.100229	0.8966286	0.5737632
P_{zi}	0.1356162	0.1849586	0.283064
DPMO_i	135,616.2	184,958.6	283,064
Z_s	0.9116183	1.0508	1.080329
P_{zs}	0.1809848	0.1466752	0.1399979
DPMO_s	180,984.8	146,675.2	139,997.9
Z_{bench}	0.4772248	0.4354062	0.1940665
P_{zbench}	0.316601	0.3316338	0.4230619
DPMO	316,601	331,633.8	423,061.9
C_{pm}	0.3338267	0.3236114	0.2672431

Tabla 4.3 Resumen de índices de capacidad por Rstudio. Elaboración propia

El índice de capacidad del proceso C_p no toma en consideración donde se localiza la media del proceso respecto de las especificaciones, simplemente mide la extensión de las especificaciones en comparación con la dispersión seis sigma del proceso, por lo tanto, podemos concluir que en las 3 líneas el proceso no es capaz y se produce un porcentaje inaceptable de productos defectuosos ya que los $C_p < 1$.

Tomando en cuenta el índice $C_{pk} = \min(C_{pi}, C_{ps})$, concluimos que si $C_p = C_{pk}$ el proceso está centrado en el punto medio de las especificaciones y cuando $C_{pk} < C_p$ el proceso esta descentrado.

En la tabla 4.4 se observa la relación entre los índices C_p y C_{pk} y se puede concluir que C_{pk} es el peor índice lateral y mientras el valor es mayor el proceso cumple mejor con las especificaciones.

Por lo tanto, las tres líneas se encuentran descentradas y las tres líneas no son capaces y requieren los procesos modificaciones serias.

Línea	C_p	C_{pk}	Conclusión
1	0.335307	0.303872	Proceso
2	0.324571	0.298876	Proceso
3	0.275682	0.191254	Proceso

Tabla 4.4 Relación Cp y Cpk. Elaboración propia

En la figura 4.1 se pueden comprender visualmente el descentrado del proceso de las tres líneas de acuerdo con las especificaciones de la empresa.

Por otro lado, el índice de Taguchi C_{pm} menciona que cumplir con las especificaciones no es sinónimo de calidad, por lo que menciona que las mejoras del proceso deben ir enfocadas en reducir la variabilidad del valor nominal en este caso 200 gramos y no solo de las especificaciones. Los valores de C_{pm} son bajos debido a que la variabilidad es alta.

Cabe mencionar que las 3 líneas no están centradas, por lo que es correcto calcular el porcentaje de paquetes defectuosos por medio de los valores de z_i y z_s que es la métrica que corresponde a la distancia entre las especificaciones y la media del proceso estos valores son diferentes entre sí ya que sirven para determinar la probabilidad de encontrar paquetes de pasta fuera de especificación por la izquierda o por la derecha y esa probabilidad sirve para poder cuantificar los DPMO (defectos por millón de oportunidad).

En la tabla 4.5 se puede observar que en la línea 1 de lado derecho se pueden encontrar más paquetes defectuosos, mientras que en la línea 2 y 3 se pueden encontrar más por la izquierda.

Línea	P(z_i)	DPMO _i	P(z_s)	DPMO _s
1	0.1356162	135,616.2	0.1809848	180,984.8
2	0.1849586	184,958.6	0.1466752	146,675.2
3	0.283064	283,064	0.1399979	139,997.9

Tabla 4.5 Probabilidad de paquetes defectuoso. Elaboración propia

4.3 Diagramas de control Xpr-R y Rango

El análisis realizado a través de histogramas, graficas de probabilidad e índices de capacidad es un resumen del desempeño del proceso; sin embargo, para analizar la capacidad potencial del proceso se tiene que abordar la cuestión del control estadístico para detectar patrones sistemáticos en la salida del proceso y poder reducir la variabilidad en cuestiones de calidad.

Como se muestra en la figura 4.3 se tienen los diagramas de control de promedio y rango para las 24 muestras de la línea 2. Los parámetros del proceso pueden estimar a partir de la carta de control como

$$\hat{\mu} = \bar{\bar{x}} = 199.37$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{16.92}{2.326} = 7.274$$

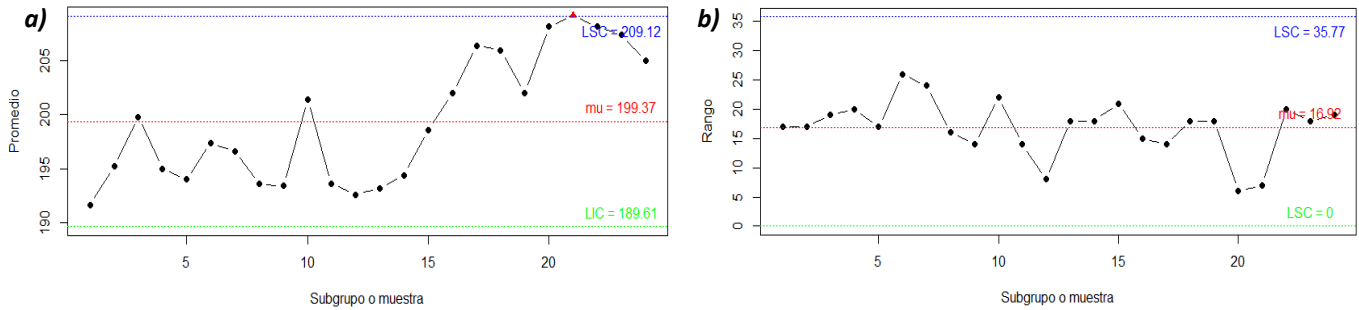


Figura 4.3 Diagramas de control a) \bar{X}_{pr} y b) R para los datos de la línea 2. Elaboración propia

tendencia central tiene un punto fuera de los límites que podría ser una causa especial del proceso.

Se espera que las medias del peso de los paquetes de pasta varíen de 189.61 a 209.12, estos límites son utilizados para detectar cambios en la media del proceso y evaluar su estabilidad.

En la figura 4.3b se ilustran cambios en la amplitud o magnitud de la variación del proceso, el diagrama de control R en donde se observa que no hay puntos fuera de los límites y además el comportamiento de los puntos no sigue ningún patrón especial; por lo tanto, el proceso responsable del peso de los paquetes de pasta ha estado funcionando de manera estable. La variación que se observa en los rangos muestrales se debe a la variación que comúnmente tiene el proceso.

Tomando en cuenta el índice de capacidad de la línea 2, $C_p = 0.324 < 1$, así que se concluye que el proceso es incapaz de cumplir especificaciones. Por lo tanto, el proceso es considerado establemente incapaz, por lo que se debe hacer análisis en busca de detectar fuentes de variación que estén en el proceso.

Como se muestra en la figura 4.4 los diagramas de control para tendencia central y amplitud o magnitud de la variación y tomando en cuenta el $C_p = 0.2756 < 1$ también es considerado establemente incapaz, por lo que también se requiere de un análisis de mejora.

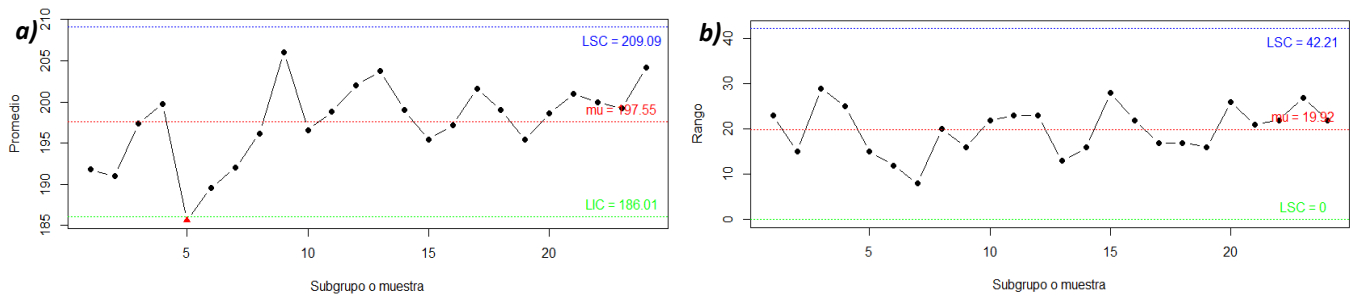


Figura 4.4 Diagramas de control a) \bar{X} y b) R para los datos de la línea 3. Elaboración propia

4.4 Diagramas de control CUSUM y EWMA

Para continuar con el análisis, se tienen los diagramas de control de suma acumulada (CUSUM) y de promedio móvil ponderado exponencialmente (EWMA) que sirven cuando se produce un desajuste muy pequeño en el proceso y por medio de estos diagramas se pueden detectar con más rapidez los cambios en la corrida de la media.

La figura 4.5 muestra el diagrama de CUSUM tabular de la línea 2 en donde se pueden ver las desviaciones de cada valor respecto del valor objetivo o nominal. Se distinguirá entre desviaciones positivas y negativas, puesto que en algunas situaciones ambas desviaciones no tienen la misma repercusión en el proceso.

En este diagrama el valor de C_i^+ y C_i^- si el proceso está en control tendrían que tender a valores nulos, en este diagrama a comparación del diagrama de \bar{X} y Rango se puede ver mejor que el proceso no se encuentra bajo control por lo que requiere que se ajuste tal vez la máquina que hace el llenado de los paquetes de pasta.

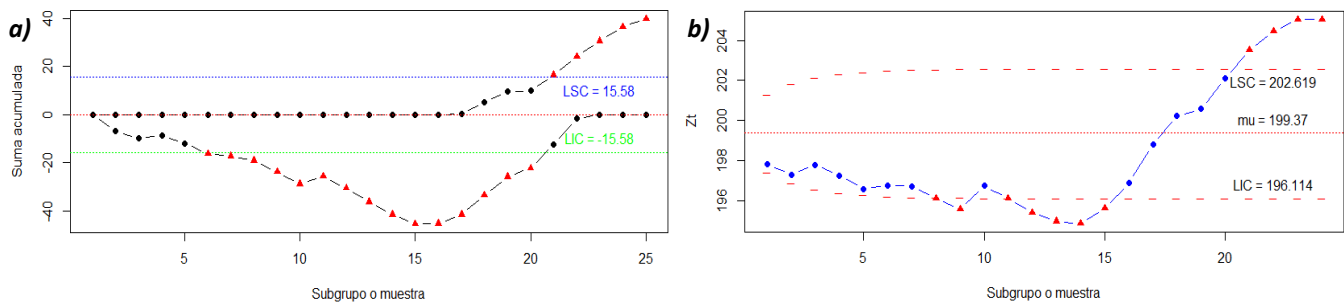


Figura 4.5 Diagramas de control a) CUSUM y b) EWMA para los datos de la línea 2. Elaboración propia

Para la línea 3 es un caso similar, debido a que en la gráfica cusum los valores de C_i^- se encuentran fuera del valor de decisión.

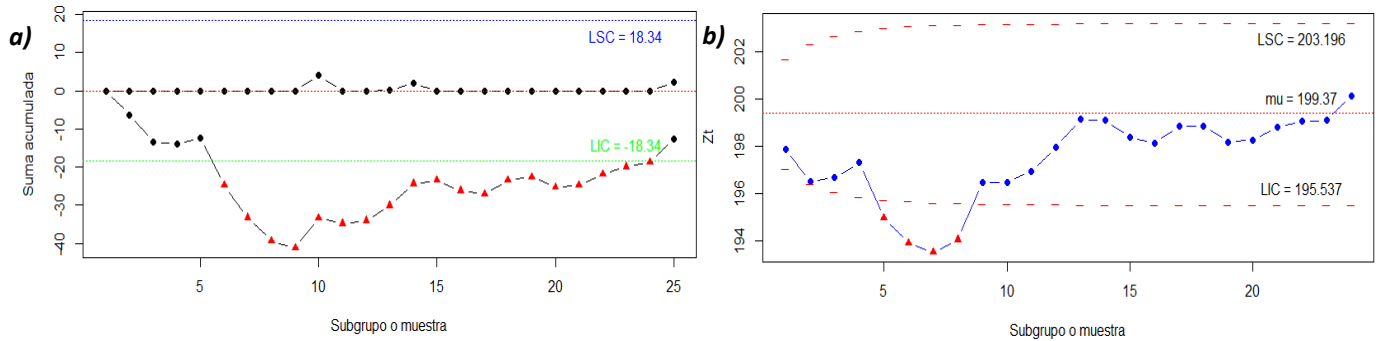


Figura 4.6 Diagramas de control a) CUSUM y b) EWMA para los datos de la línea 3. Elaboración propia

4.5 Diagramas de control empleando la distribución binomial con n variable

En la figura 4.7a se ilustra el diagrama de control para fracción disconforme con subgrupos variables, pero se utilizó el tamaño de subgrupo promedio \bar{n} , lo cual representa que

$$\bar{n} = \frac{\text{Total inspeccionados}}{\text{Total de subgrupos}} = \frac{1440}{24} = 60$$

La interpretación de este diagrama de control es que de cada 60 paquetes se espera que la proporción de paquetes con problemas de peso varíe de 0.0 a 0.1, con un promedio de 0.03. Al multiplicar por 100, se puede interpretar en términos porcentuales de la siguiente manera: se espera que el porcentaje de paquetes con problemas de peso varíe de 0.0 y 10% con un promedio de 3%.

De acuerdo con lo anterior, en la gráfica 4.7a se tiene un punto fuera del límite superior de control, por lo cual se concluye que el proceso no fue estable, y se puede decir que el proceso funcionó en presencia de una causa especial que por lo general no está presente en el proceso, y que causó que la proporción de disconformes fuera anormalmente grande.

Por lo tanto, a pesar de que el proceso tiene solo un punto fuera del límite superior de control, su desempeño se encuentra lejos de ser satisfactorio ya que el promedio de defectos es relativamente alto 3%. Por lo que se propone revisar el proceso para encontrar las causas comunes que están ocasionando el problema.

En la figura 4.7b se tiene el diagrama de control p normalizado, que en lugar de graficar la proporción muestral p_i , se grafica la proporción estandarizada.

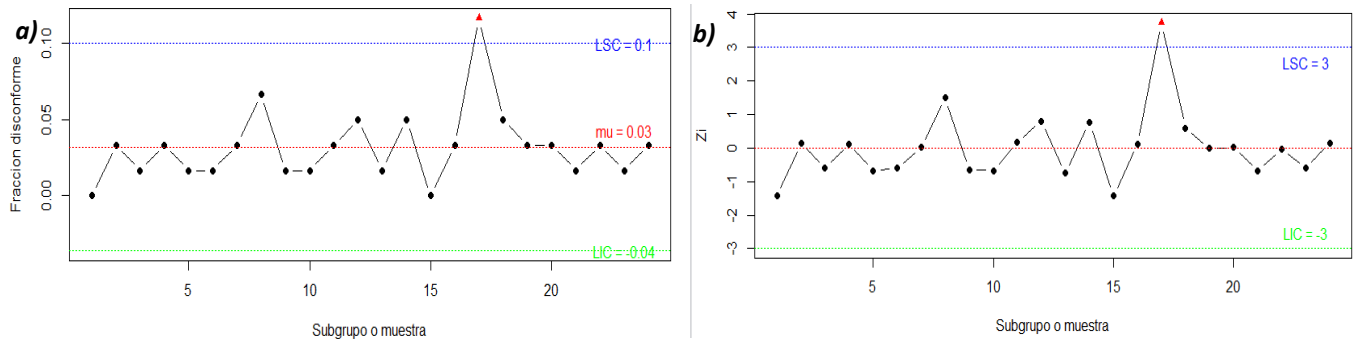


Figura 4.7 Diagrama de control a) fracción disconforme (p) b) *normalizada* para los datos de la línea 2. Elaboración propia

En la figura 4.8a se muestra la línea 3 no tiene ningún punto fuera de los límites de control, pero por cada 60 paquetes de pasta se espera que la proporción de paquetes con problemas de peso varíe de 0.0 a 0.12, con un promedio de 0.04. En términos porcentuales se espera que el porcentaje de paquetes con problemas de peso varíe de 0.0 y 12% con un promedio de 4%, por lo cual requiere de una implementación de mejora para disminuir el problema. En la figura 4.8b se tienen el diagrama p normalizado.

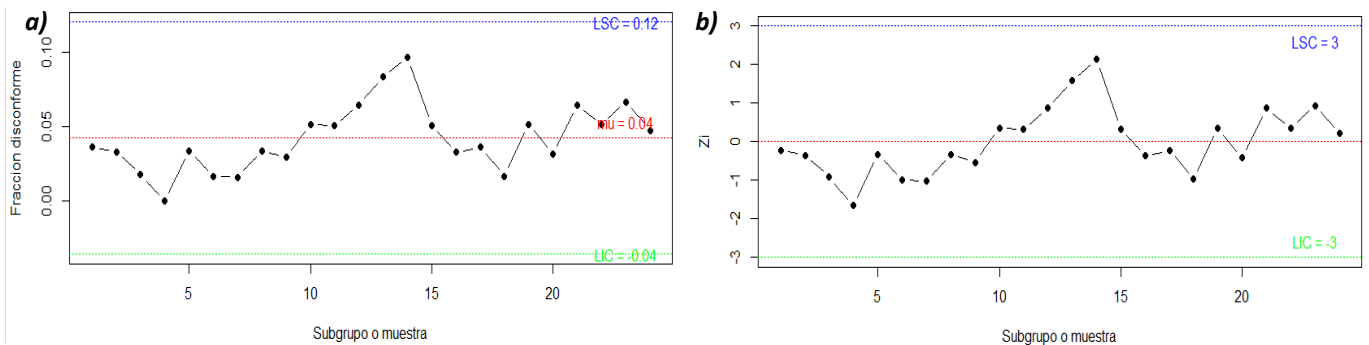


Figura 4.8 Diagrama de control a) fracción disconforme (p) b) *normalizada* para los datos de la línea 3. Elaboración propia

4.6 Diagramas de control empleando la distribución de Poisson

El diagrama de control u analiza la variación del número promedio de defectos por artículo o unidad de referencia, debido a que el tamaño de subgrupo no es constante se utiliza

$$\bar{n} = \frac{\text{Total de defectos}}{\text{Total de artículos inspeccionados}} = \frac{152}{1440} = 0.1055$$

En la figura 4.9a se pueden ver dos puntos fuera de los límites de control superior lo que demuestra que el proceso no trabaja de manera estable, y que por cada paquete de pasta se encuentran de 0.0 a 0.23 defectos con un promedio de 0.11.

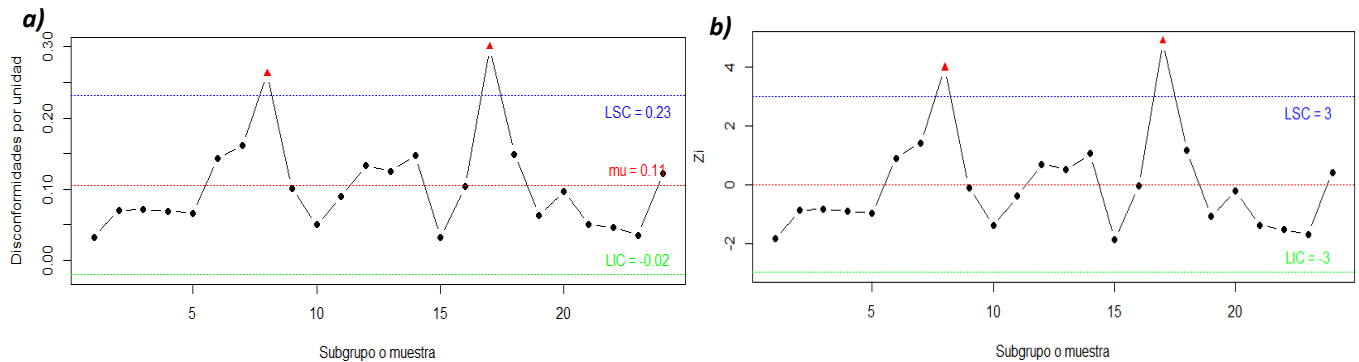


Figura 4.9 Diagrama de control a) disconformidades por unidad (u) b) normalizada para los datos de la línea 2. Elaboración propia

La figura 4.9b muestra un diagrama de control estandarizado en caso de que los tamaños de subgrupo sean muy distintos entre sí, en la cual, en lugar de graficar u_i se grafica la siguiente variable estandarizada.

Para el análisis de la línea 3, en la figura 4.10 no se tiene ningún punto fuera de los límites de control, pero por cada paquete de pasta se encuentran de 0.0 a 0.29 defectos con un promedio de 0.14.

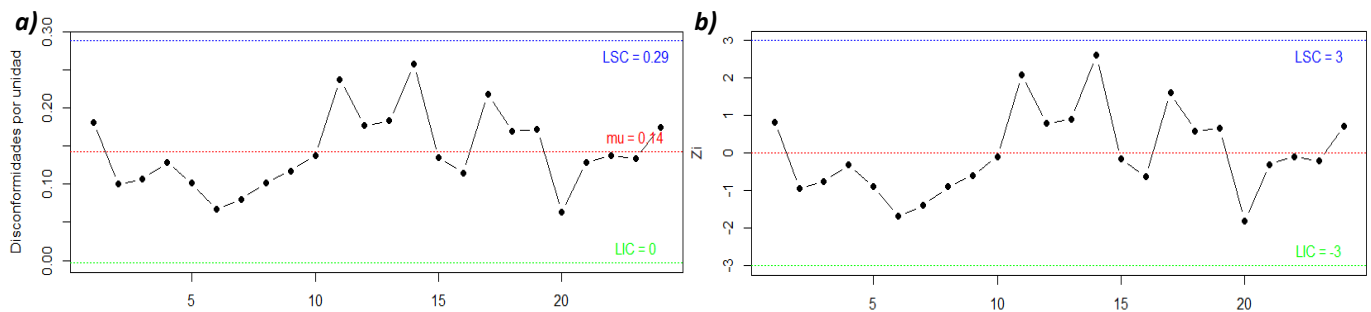


Figura 4.10 Diagrama de control a) disconformidades por unidad (u) b) normalizada para los datos de la línea 3. Elaboración propia

Por último, todo el análisis que se hizo con Rstudio fue con base en los datos de la línea 1; sin embargo, los resultados obtenidos también se pueden conseguir con ayuda de excel y minitab y por eso se ha elaborado un análisis de la línea 2 con minitab que se encuentra en la sección de anexos en el punto 7.1 y un análisis de la línea 3 con ayuda de excel en el punto 7.2 en donde se ven reflejados los resultados del análisis pero en programas diferentes.

5 Conclusiones

Durante la elaboración de este trabajo fue posible desarrollar los códigos en Rstudio sin tener conocimiento del software y así poder mostrar su aplicabilidad para estudiantes de la carrera de Ingeniería Industrial en el área de la estadística, específicamente en el análisis estadístico de control de procesos.

Al iniciar este trabajo el conocimiento sobre el software era casi nulo, así como también su funcionalidad y la lógica que seguía el lenguaje; sin embargo, el lenguaje resultó ser intuitivo y, para ser lenguaje de programación resultó ser amigable, por lo que el aprendizaje de su estructura fue rápido. Uno de los puntos importantes fue que no es necesario tener un conocimiento tan amplio en programación ya que el software te lleva de la mano. La existencia de una gran cantidad de blogs en internet ayuda a resolver dudas que se te van presentando durante la elaboración del código. La gran cantidad de materiales disponibles es debido a que es un software libre.

Cabe mencionar que al realizar personalmente el código ayudo a comprender mejor los temas relacionados con el control estadístico de proceso, que sirve para mejorar la calidad en los productos o servicios, porque a pesar de que se trabaja con fórmulas estandarizadas al obtener el resultado de los datos, se tiene conocimiento de cómo se obtuvo y el por qué se obtiene ese resultado, en comparación con otros paquetes, por ejemplo minitab por mencionar alguno, es excelente en el análisis de datos estadísticos de calidad; sin embargo, solo plasman un resultado a diferencia de Rstudio que ese resultado se obtiene de un código elaborado a la medida del análisis que se busca obtener.

Una de las ventajas que tiene Rstudio en comparación con excel es que es menos tardado hacer un análisis similar, en Rstudio es muy útil y cómodo usarlo para trabajos repetitivos y es poco probable que te equivoques por error humano. Si tu código y/o procedimiento de cálculo está bien para un dato o conjunto de datos, está bien para todos. En excel se tendría que repetir los cálculos para cada proyecto o grupo de datos esa es una de las ventajas que tiene este software.

Otra de las ventajas es que la calidad de gráficos en Rstudio es muy amplia y se puede editar de acuerdo con las necesidades de cada usuario, a diferencia de minitab que los

gráficos tienen un formato predeterminado. En excel los gráficos son también de acuerdo con la necesidad del usuario, pero no hay tantas opciones de editar la figura, por ejemplo, graficar las distribuciones de probabilidad es difícil y graficar la distribución de probabilidad encima de un histograma es aún más complejo que con ayuda de Rstudio este tipo de gráficos es rápido de hacer.

Cabe mencionar que durante la investigación se concluyó que no es el paquete más usado ya que actualmente es excel, ni el más rápido porque es C++, pero es un buen intermedio para llevar a cabo un buen análisis de datos, una de las mejores formas de medir la popularidad en el mercado del software para la ciencia de datos es contar el número de anuncios de trabajo para cada uno y cada vez es más común encontrar empleos que solicitan que se tenga conocimiento sobre el proceso de información en R, esto quiere decir que las tendencias laborales dan una buena idea de lo que probablemente se volverá popular en el futuro. Por lo cual se vuelve relevante que en las universidades se haga conocimiento de la importancia que tiene este software y la tendencia va incrementando al alza.

Por lo tanto, R no es la única opción para utilizar y tal vez no el más conveniente, todo depende del uso que quieras darles a tus datos, así como el tamaño de estos, existen muchos más softwares, pero el motivo de este trabajo es dar a conocer a los estudiantes de la carrera de Ingeniería Industrial como herramienta potente de análisis de datos, pero si sería muy conveniente que los egresados de la FI lo conozcan y que no le tengan miedo porque es programación, como al inicio de este trabajo se pensaba; sin embargo, queda demostrado que el procesos durante la elaboración del código no fue complicado y es recomendable.

Por último, vale la pena mencionar que puedes mezclar paquetes ya elaborados por algunos otros estudiantes o profesores alrededor del mundo, por ejemplo, sobre finanzas, pronósticos, six sigma y al mismo tiempo ocupar paquetes para perfeccionar tus gráficos que son importantes para que la visualización de tu análisis sea más amigable durante alguna presentación. Realizando una pequeña comparación con minitab el cual es un software potente en estadística, pero solo se aplica para estadística y no para temas laterales a diferencia de R que se puede usar para muchas otras aplicaciones de análisis de datos y big data.

6 Bibliografía

- Allen, W. (2001). *Estadística aplicada a los negocios y la economía. 3a Edición*. México: McGraw Hill.
- Correa, J. C., & Nelfi, G. (2002). *Graficos estadísticos con R*. Obtenido de <https://cran.r-project.org/doc/contrib/grafi3.pdf>
- GNU. (5 de Mayo de 2016). *¿Qué es un software libre?* Recuperado el 11 de Junio de 2016, de GNU: <http://www.gnu.org/>
- Gutiérrez Pulido, H., & De la Vera Salazar, R. (2009). *Control estadístico de la calidad y seis sigma*. México D.F.: McGraw-Hill.
- Jose Miguel Contreras Garcia, E. M. (2010). *Introduccion a la programacion estadística con R para profesores*. Granada: Los autores.
- Juran M., J. (1990). *Juran y el liderazgo para la calidad*. Ediciones Diaz de Santos, S.A.
- Leo Revilla, Á. (7 de Agosto de 2013). *Breve historia de R*. Recuperado el 11 de Junio de 2016, de Momentotic: <https://momentotic.com/2013/08/07/breve-historia-de-r/>
- Montgomery, D. C. (2016). *Control Estadístico de la Calidad* (Tercera ed.). Mexico: Limusa.
- Paradis, E. (2003). *R para principiantes*. Recuperado el 3 de Junio de 2017, de https://cran.r-project.org/doc/contrib/rdebut_es.pdf
- Purnal, A. W. (2012). *Análisis de la capacidad de procesos industriales*. Mexico: Facultad de Ingeniería, Universidad Nacional Autónoma de México.
- Ross, I. (s.f.). *R : Past and Future History*. Recuperado el 11 de Junio de 2016, de The University of Auckland: <https://www.stat.auckland.ac.nz/~ihaka/downloads/Interface98.pdf>
- The R Project. (s.f.). *¿What is R?* Recuperado el 11 de Junio de 2016, de The R Project for Statistical Computing: <https://www.r-project.org/>
- Vance, A. (7 de Enero de 2009). Data Analysts Captivated by R's Power. *New York Times*, pág. B6
- Verzani, J. (2011). *Getting started with RStudio*. United State of America: O'Reilly Media.

7 Anexos

7.1 Minitab

Resultados para la Línea 2.

Variable	N	N*	Media	Error estándar de la media	Desv.Est.	Mínimo	Q1	Mediana	Q3	Máximo
Datos	120	0	199.37	0.788	8.63	185.00	192.00	198.00	206.00	215.00

Figura 7.1 Estadísticos descriptivos. Elaboración propia

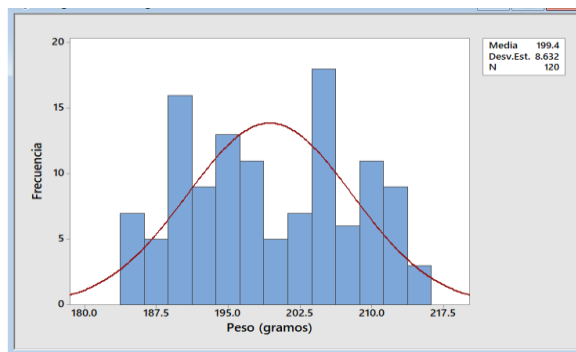


Figura 7.2 Histograma. Elaboración propia

Prueba de $\mu = 200$ vs. $\neq 200$
La desviación estándar supuesta = 8.632

Variable	N	Media	Desv.Est.	Error estándar de la media	IC de 95%	Z	P
Peso (gramos)	120	199.367	8.632	0.788	(197.822, 200.911)	-0.80	0.422

Figura 7.3 prueba de hipótesis (Z de una muestra). Elaboración propia

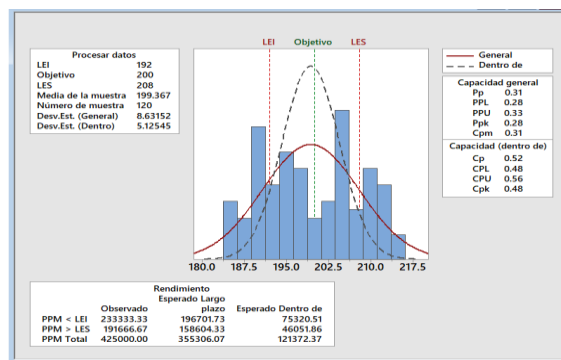


Figura 7.4 Informe de capacidad del proceso. Elaboración propia

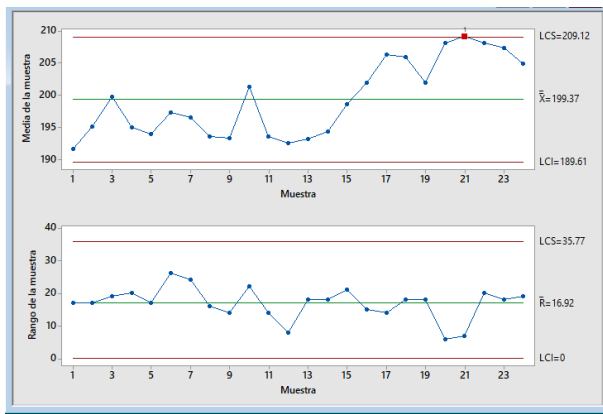


Figura 7.5 Diagrama de control Xbarra-R. Elaboración propia

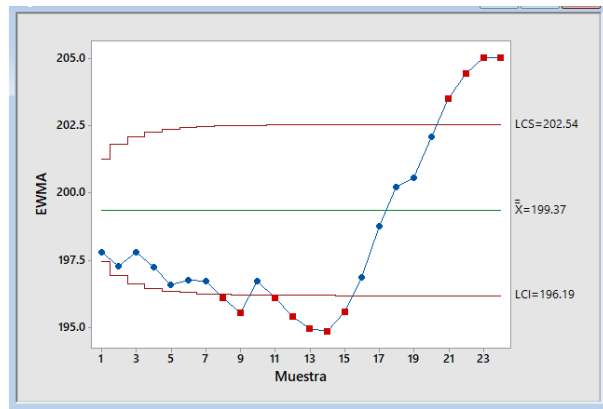


Figura 7.6 Diagrama de control de tiempo ponderado EWMA. Elaboración propia

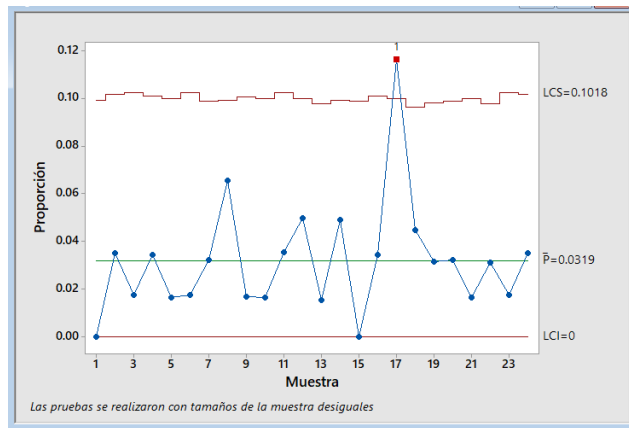


Figura 7.7 Diagrama de control fracción disconforme (p). Elaboración propia

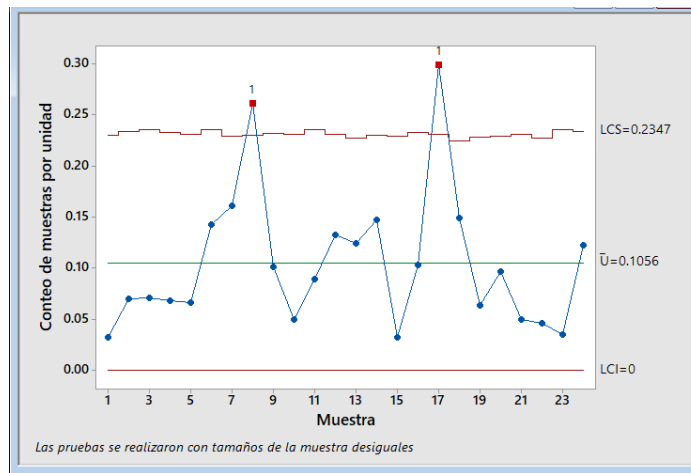


Figura 7.8 Diagrama de control disconformidades por unidad (u). Elaboración propia

7.2 Excel

Resultados para la Línea 3

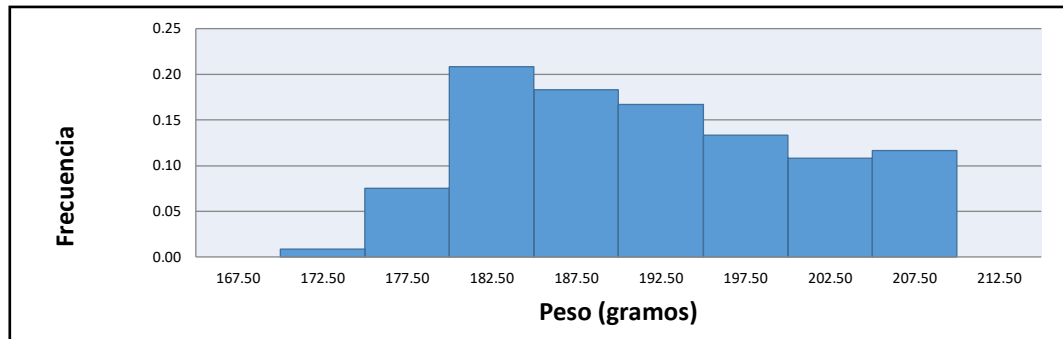


Figura 7.10 Histograma. Elaboración propia

Media	197.55
Mediana	197
Máximo	215
Mínimo	175
Q1	190
Q3	205
IQR	15
Desviación	8.988
Varianza	80.11

Tabla 7.9 Estadísticos descriptivos. Elaboración propia

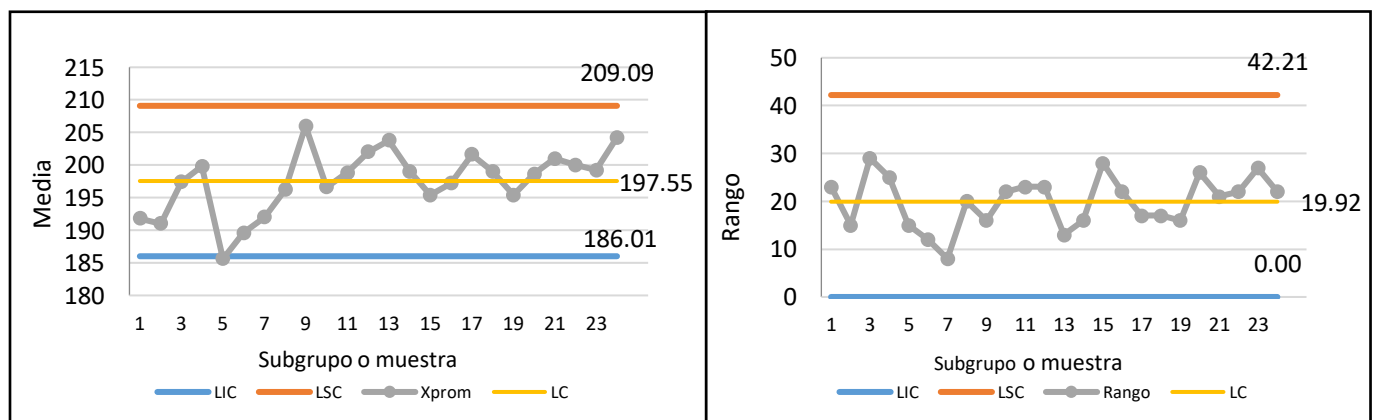


Figura 7.11 Diagrama de control Xbarra y R. Elaboración propia

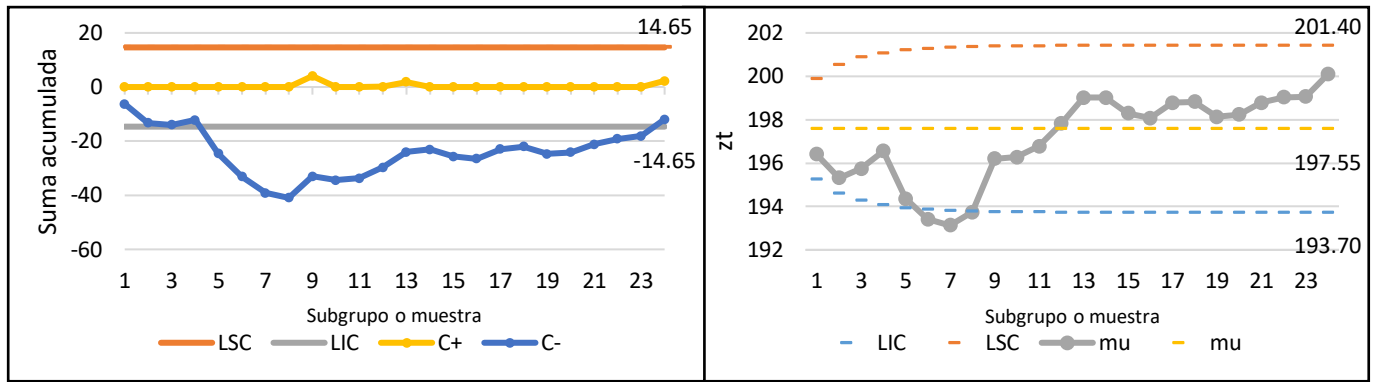


Figura 7.12 Diagrama de control EWMA y CUSUM. Elaboración propia

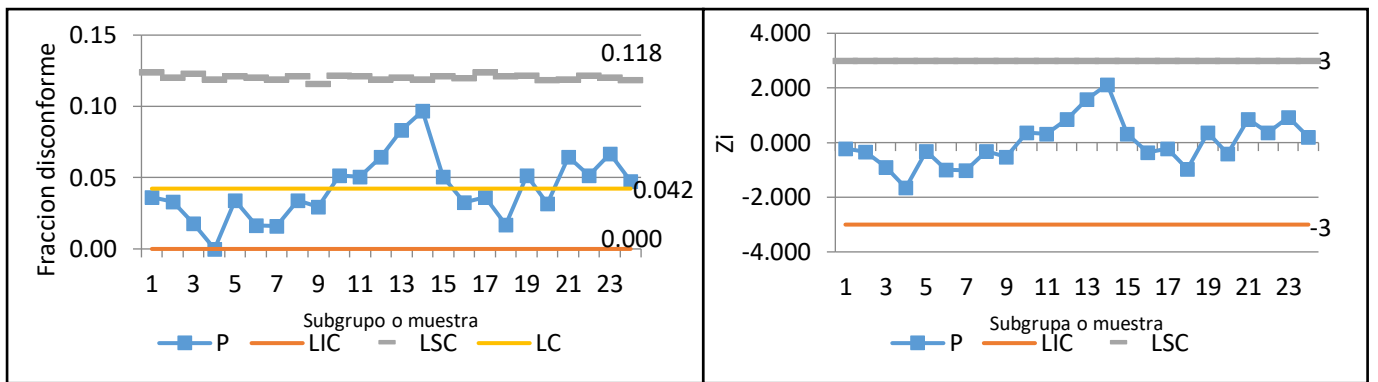


Figura 7.10 Diagrama de control a) Fracción disconforme (p) b) Normalizada. Elaboración propia

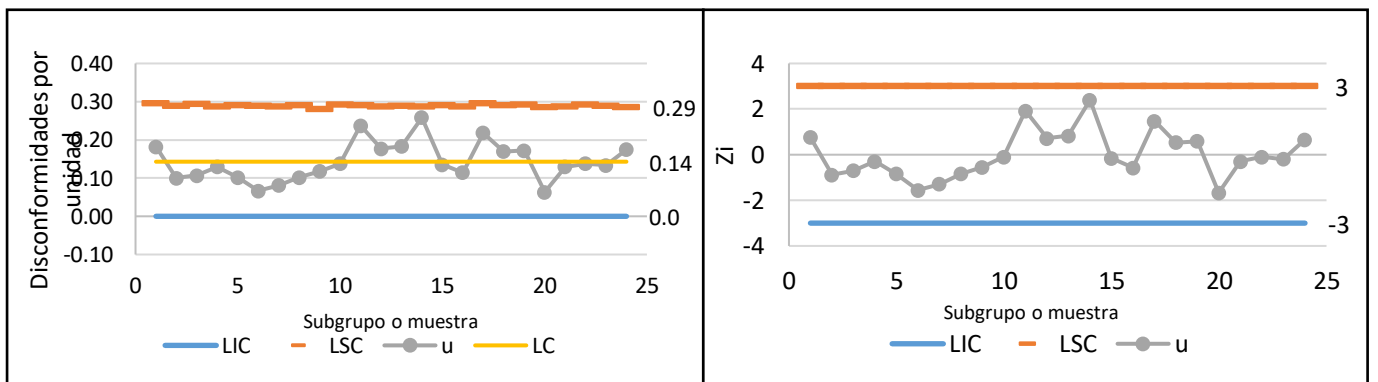


Figura 7.14 Diagrama de control a) Disconformidades por unidad (u) b) normalizada

LIE =	192		
LSE =	208		
LSE - LIE =	16		
N =	200		
	R_{pr}/d₂	S_{pr}/C₄	√(s²_{pr})
	8.600	8.622	8.461
C _p =	0.3101	0.3093	0.3152
¿Capaz?	NO	NO	NO
C _r =	3.2248	3.2331	3.1728
C _{pi} =	0.2151	0.2146	0.2187
C _{ps} =	0.4051	0.4040	0.4117
C _{pk} =	0.2151	0.2146	0.2187
z _i =	0.65	0.64	0.66
P(z < z _i) =	0.25934	0.25987	0.25592
DPMO _i =	259341	259874	255920
z _s =	1.22	1.21	1.24
P(z > z _s) =	0.1122	0.1127	0.1084
DPMO _s =	112150	112742	108392
z _{bench} =	0.328	0.325	0.347
1 - P(z _i < z < z _s) =	37.15%	37.26%	36.43%
DPMO =	371491	372616	364313
P(z _i < z < z _s) =	63%	63%	64%
t =	8.9418	8.9629	8.8083
C _{pm} =	0.2982	0.2975	0.3027

Tabla 7.1 Índices de capacidad línea 3. Elaboración propia