UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTA
CONTROL ESTADISTICO DE CALIDAD
ACTIVIDAD EN CLASE

Cesar Augusto Prieto Sarmiento - CC: 1065843742
Cristian Camilo Prieto Zambrano - CC: 1001316206

14 de marzo de 2024

PUNTO 1

En la tabla que se muestra a continuación de este enunciado, se reportan los tres últimos dígitos de las mediciones de los diámetros interiores de un cilindro para la construcción de los motores de cierta marca comercial de automóvil. El régimen de la producción de los cilindros es tal que las muestras se pueden recolectar cada media hora, pero con tamaños de máximo cinco unidades. Es de interés establecer si el proceso se encontraba bajo control estadístico cuando se recolectaron las muestras, mediante el diseño de las Cartas $\overline{\boldsymbol{X}}$ y $\boldsymbol{R}$.

Para cada una de las muestras, encontramos el tamaño de la misma, la media y la varianza. Estos datos nos ayudarán en la construcción de las LC (Lineas Centrales) de $\bar{\bar{X}}$ y $\bar{S}$

MUESTRA	x_1	x_2	x_3	x_4	x_5
1	205	202	204		
2	207	205	202		
3	196	201	198	202	
4	203	198	196	217	
5	201	202	199		
6	197	203			
7	205	196	201		
8	197	199	196		
9	201	200			
10	195	203	204	199	200
11	202	202			
12	198	203			
13	202	196	200		
14	201	187	209	202	200
15	202	196	204	195	197
16	200	204	197	199	
17	197	199	201	201	
18	205	204	202	200	
19	200	201	199	200	
20	201	205	196	201	
21	197	198	199		
22	200	200	201	205	201
23	202	202	204		
24	198	203	201	198	
25	204	201	201		
26	206	194	197		
27	200	204	198		
28	199	199			
29	198	204			
30	203	200	204	199	200
31	196	203	197	201	
32	197	199	203		
33	197	194	199	200	199
34	203	201	196	201	

Como estamos trabajando con diferentes tamaños de muestra para las 34 presentes en la tabla, debemos trabajar con cartas de control diseñadas para muestras variables. Encontramos la construcción de las lineas centrales para la carta $\bar{\bar{X}}$ y para $\bar{S}$, en donde ponderamos por cada uno de los tamaños de las muestras.

Para la linea central de $\bar{X}$ tenemos lo siguiente

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{m} n_i \bar{x}_i}{\sum_{i=1}^{m} n_i}$$

Y obtenemos como estimación de $\bar{\bar{X}}$ a 200.4067

Para la linea central de $\bar{S}$ tenemos lo siguiente

$$\bar{S} = \left[\frac{\sum_{i=1}^{m} (n_i - 1) s_i^2}{\sum_{i=1}^{m} n_i - m} \right]^{1/2}$$

Y obtenemos como estimación de $\bar{S}$ a 3.703200

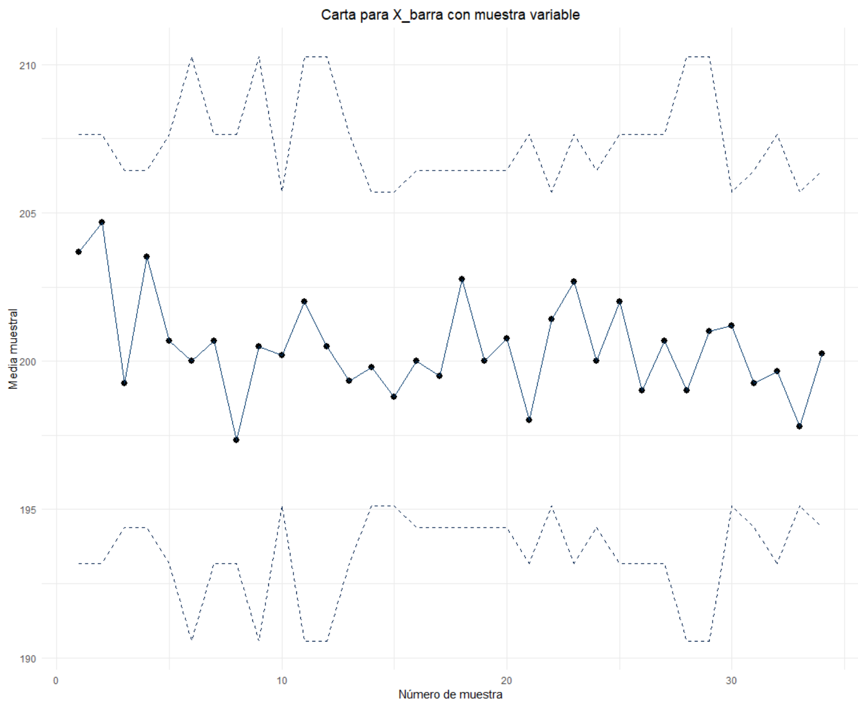
Solucion:
Contruímos la Carta para $\bar{\bar{X}}$ con límites de control dados por:

$$\begin{gathered} \text{UCL} = \bar{\bar{X}} + A_3 \bar{s} \\ \text{CL} = 200.4067 \\ \text{LCL} = \bar{\bar{X}} - A_3 \bar{s} \end{gathered}$$

En donde A_3 es una constante que se encuentra tabulada para un n fijo, en este caso, como el tamaño de la muestra se encuentra variando de 2 a 5, tenemos los siguientes valores para A:

$A_3 = (2{,}659, 1{,}954, 1{,}628, 1{,}427)$ para n_2, n_3, n_4, n_5 respectivamente.

Sabiendo esto y contando con los datos, realizamos la gráfica de la Carta de control para $\bar{\bar{X}}$ con muestra variable



Podemos evidenciar en la gráfica que ninguna de las medias muestrales se ubica por fuera de los límites de control, además, basandonos en las reglas de sensibilización podemos decir que no se evidencia ningún patrón aparente en la misma, por lo que concluimos que el modelo se encontraba en control para el momento en que fueron tomadas las muestras.

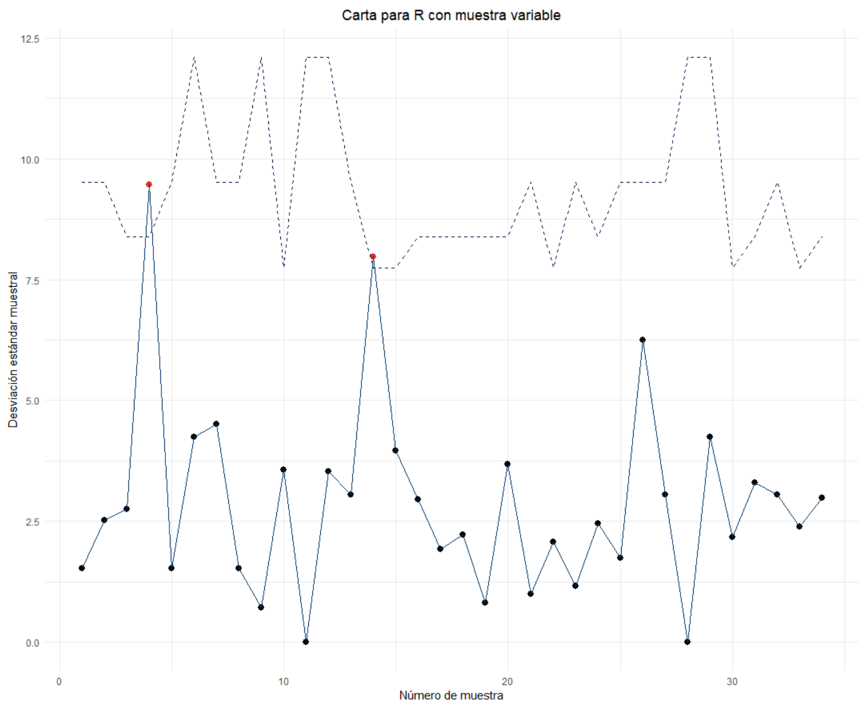
Solucion:
Contruimos la carta R como lo haríamos para una carta S en Fase 1, de manera que los límites de control se contruyen con los valores de B_3 y B_4 que se encuentran tabulados, en donde podemos evidenciar que los valores para B_3 son 0 para muestras de tamaño 2,3,4 y 5, por lo que no tendremos en cuenta el límite inferior el cual nos indicaría una disminución en la dispersión de los datos, lo que en general no es malo. Lo que nos va a interesar será saber que tanto aumenta la dispersión de nuestros datos.
Construimos el la carta de control S de la siguiente manera:

$$\text{UCL} = B_4\bar{s}$$
$$\text{CL} = 3.703200$$

Para valores de B_4 dados por:

$B_4 = (3.267, 2.568, 2.266, 2.089)$ para n2, n3, n4, n5 respectivamente.

Sabiendo esto y contando con los datos, realizamos la gráfica de la Carta de control para R con muestra variable



Podemos evidenciar que el proceso no se encuentra bajo control dado que la muestra 4 y la muestra 14 se encuentran por fuera de los límites de control, lo que nos indica que la dispersion para estas muetras es mayor de lo estipulado y habría que verificar las razones por las que el proceso se salió de control, en caso de que no estén ligadas a la variación inherente del mismo.

PUNTO 2

Sea $X \sim N(\mu; \sigma)$ una variable de calidad. Se sabe que $\mu_0 = 20$ y $\sigma_0 = 3$ son, respectivamente, los valores objetivo de la media y la desviación estándar del proceso. Para metodología 6-sigma, con subgrupos racionales de tamaño $n = 5$. Se presenta a continuación una longitud de corrida del proceso:

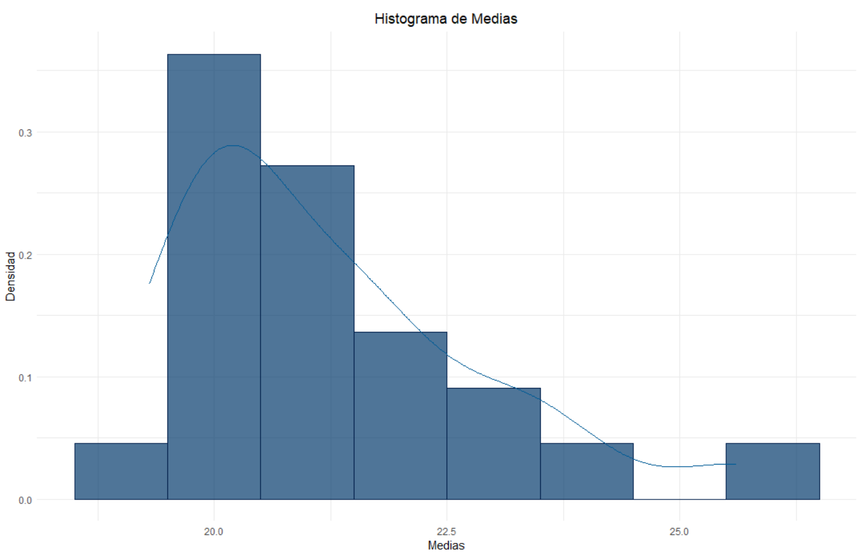
MUESTRA	X_1	X_2	X_3	X_4	X_5
1	17,067	20,761	19,547	21,635	18,887
2	25,413	15,350	19,811	23,132	17,232
3	22,548	20,635	20,421	19,063	16,663
4	19,282	23,362	19,569	24,033	17,397
5	20,953	19,295	23,526	14,093	18,652
6	21,593	15,979	24,727	19,421	16,318
7	18,365	18,910	16,433	21,288	25,576
8	19,493	22,219	20,231	20,299	18,602
9	13,524	24,450	18,814	20,996	21,113
10	16,927	19,129	22,343	21,699	22,781
11	19,049	25,734	23,490	22,559	25,448
12	17,802	19,587	19,089	26,604	19,432
13	20,193	21,247	21,452	23,967	19,858
14	20,641	26,829	24,511	24,335	23,155
15	25,706	24,212	23,940	24,316	29,853
16	17,833	21,487	28,196	22,616	21,437
17	22,565	20,573	22,077	18,646	24,293
18	21,771	22,152	21,579	22,612	22,274
19	18,783	24,608	24,360	18,249	20,992
20	25,248	17,629	20,237	18,170	26,116
21	20,361	17,164	20,477	23,651	19,669
22	20,432	26,040	21,984	22,831	24,316

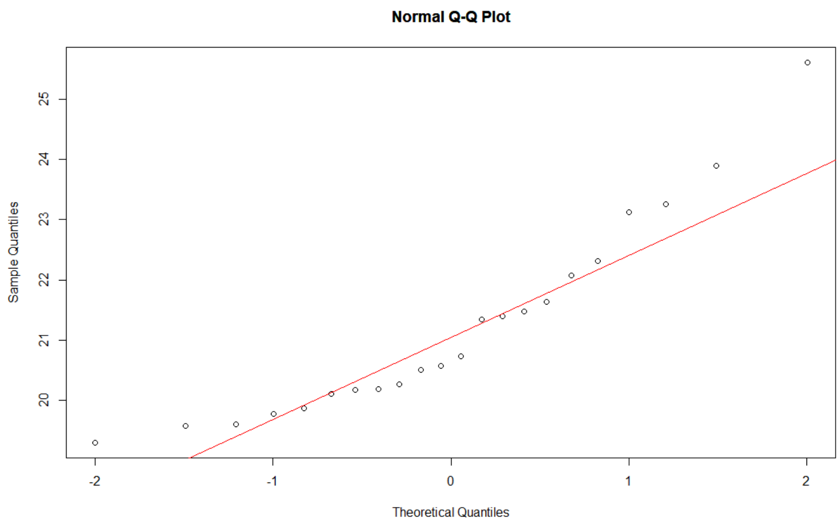
Se pide:

1. **Establecer si el proceso es estable con respecto al nivel medio.**

   Solucion:

   Primero comprobemos hipotesis de normalidad, empezando con un histograma de la densidad de las medias obtenidas en cada muestra y un qqplot, para explorar tambien la distribucion teorica con respecto a los cuantiles de la media.



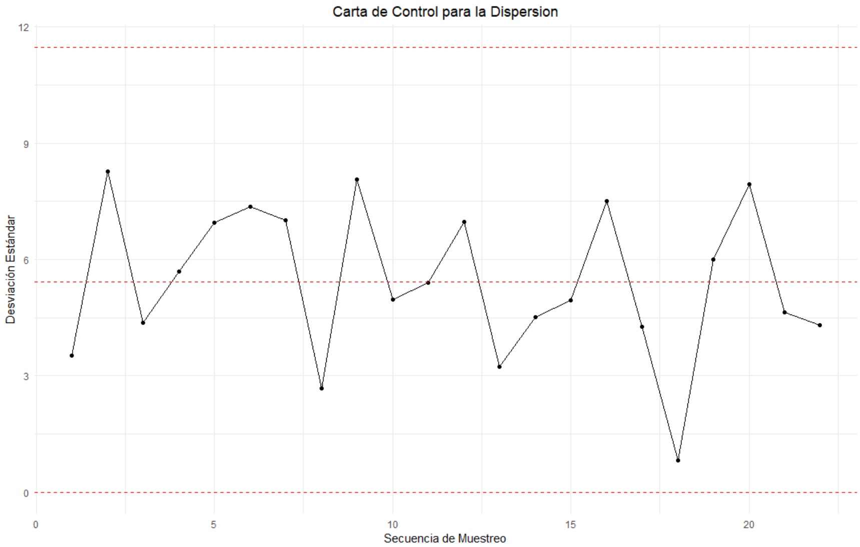


Las graficas nos dan una idea muy vaga de loa distribucion que pueden llegar a tener nuestros datos, así que, por otra parte, tambien realizamos la prueba Chi cuadrado de normalidad de Pearson la cual nos dio como resultado:

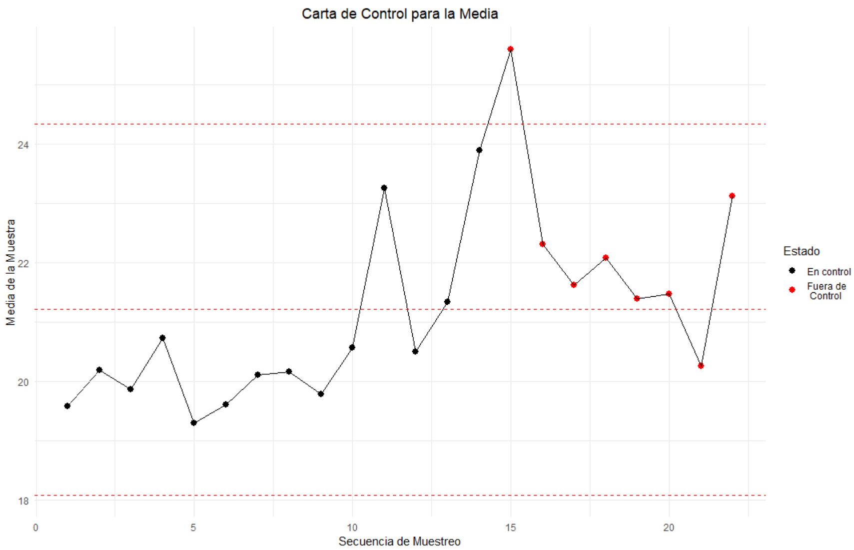
```
        pearson.test(B)

    Pearson chi-square normality test

data:  B
P = 7.0909, p-value = 0.7917
```

Y ya con esto podemos decir que dado que el valor p es mayor que cualquier nivel de significancia comúnmente utilizado (como 0.05), no hay suficiente evidencia para rechazar la hipótesis nula. Por lo tanto, en base a este resultado, podríamos concluir que no hay suficiente evidencia para decir que los datos no provienen de una distribución normal.

Con esto procedemos con la creacion de nuestras cartas de control utilizando el metodo de 6-sigma para los intervarlos o limites de confianza que usaremos para el monitoreo del comportamiento de nuesta variable caracteristica de interes, luego de realizar los calculos para la carta S obtenemos la siguiente grafica:



Como no notamos ninguna anomalia, es decir, no vemos que ningun punto se sale de control con respecto a la desviacion, entonces procedemos a realizar la carta para $\bar{X}$, la cual quedaria como se muestra a continuacion:



En la grafica anterior podemos notar que el punto 14 muestra una subida brusca en la carta de control lo que tal vez ocaciona que el punto 15 ya se muestre por fuera de los limites de control establecido a travez de los calculos realizados, en este punto el proceso se salio de control con respecto a $\bar{X}$

2. **Si el proceso se ha salido de control, establézcase aproximadamente el momento en el que ocurrió la salida y la magnitud del cambio ocurrido.**

   Solucion: Como conclucion del punto anterior, partimos desde el hecho que la carta para la caracteristica de interes estudiada con respecto a $\bar{X}$ se ha salido de control entonces procedemos a identificar la magnitud del cambio a travez de la diferencia de medias, con lo que obtenemos lo siguiente:

```
# Calcular la magnitud del cambio
media_antes <- mean(AA1$Medias[1:(punto_fuera_de_control - 1)])
```

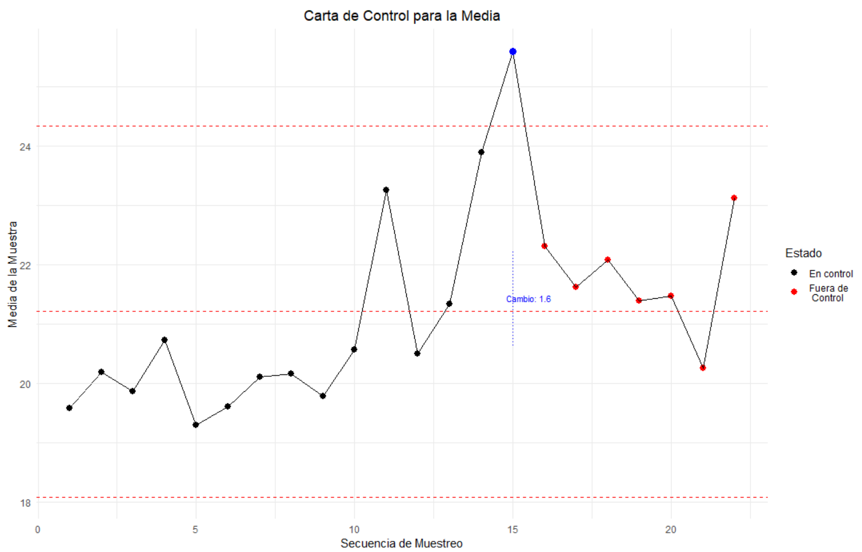
```
media_despues <- mean(AA1$Medias[punto_fuera_de_control:nrow(AA1)])
magnitud_cambio <- abs(media_despues - media_antes)

# Imprimir resultados
cat("El proceso se salió de control aproximadamente en la muestra:", punto_fuer
```

```
El proceso se salió de control aproximadamente en la muestra: 15
```

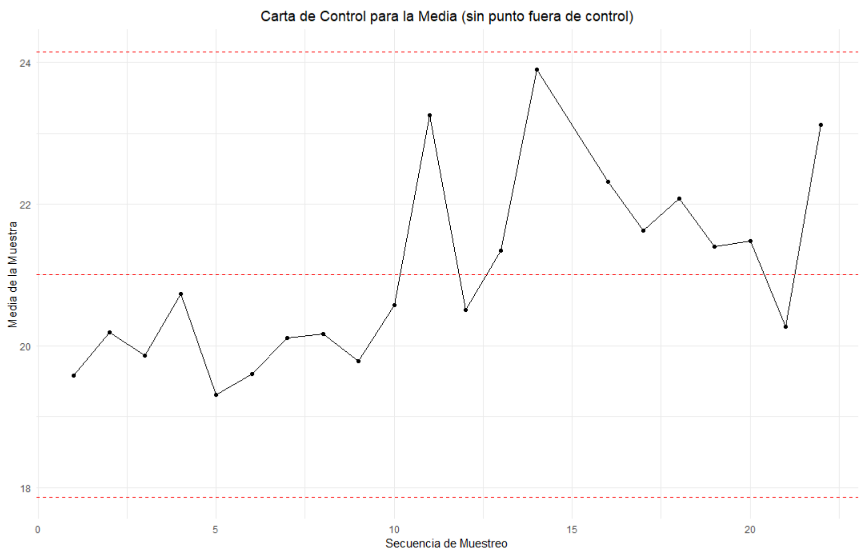
```
cat("La magnitud del cambio fue de:", magnitud_cambio, "\n")
```

```
La magnitud del cambio fue de: 1.600104
```

De esta forma, como ya conociamos anteriormente, el punto donde el proceso se salio de control fue en la muestra 15 y la magnitud del cambio fue aproximadamente de 1.6 unidades, esto lo podemos evidenciar más claramente en el siguiente grafico:



Ahora, como extra al ejercicio, la siguiente imagen muestra como se ve la carta sin incluir la muestra que se salio de control.



Aunque seguimos viendo puntos muy cercanos al limite superior de control, estos no llegan a revasarlo y tampoco a tocarlo, entonces podríamos concluir que sin tener en cuenta la muestra que se salio de control el proceso ha permanecido en control durante cada observacion de la muestra.

CODIGO

```
#################################################################################
################################ CODIGO PUNTO 1 #################################

Datos6 <- read_excel("Datos6.xlsx", col_types = c("skip","numeric", "numeric", "numeric"

# Función para calcular la media ponderada de un conjunto de datos
# Parámetros:
#   - data: Conjunto de datos
#   - na.rm: Booleano que indica si se deben omitir los valores NA (por defecto es TRUE)
# Salida:
#   - Media ponderada del conjunto de datos
calcular_media_ponderada <- function(data, na.rm = TRUE) {
  medias_por_fila <- rowMeans(data, na.rm = na.rm)  # Calcula la media de cada fila omit
  suma_ponderada <- sum(medias_por_fila * rowSums(!is.na(data)))  # Suma de las medias p
  suma_observaciones <- sum(rowSums(!is.na(data)))  # Suma de los números de observacion
  return(suma_ponderada / suma_observaciones)  # Retorna la media ponderada
}

# Calcular el tamaño de la muestra
tamaño_muestra <- rowSums(!is.na(Datos6))

# Calcular la media de cada muestra
media_muestra <- rowMeans(Datos6, na.rm = TRUE)

# Calcular la varianza de cada muestra
varianza_muestra <- apply(Datos6, 1, function(x) var(x, na.rm = TRUE))

# Crear un nuevo dataframe con los resultados
nuevo_database <- data.frame(Tamaño_Muestra = tamaño_muestra,
                             Media_Muestra = media_muestra,
                             Varianza_Muestra = varianza_muestra)

# Imprimir el nuevo dataframe
print(nuevo_database)

# Calcular la media ponderada de todo el conjunto de datos
resultado_final <- calcular_media_ponderada(Datos6)

# Imprimir el resultado final
print(paste("Resultado final:", resultado_final))

#-----------------------------------------------------------------

# Función para calcular el LC_Sbarra
```

```
# Parámetros:
#   - tamaño_muestra: Vector con los tamaños de muestra
#   - varianza_muestra: Vector con las varianzas de muestra
# Salida:
#   - LC_Sbarra: Valor de LC_Sbarra
calcular_LC_Sbarra <- function(tamaño_muestra, varianza_muestra) {
  numerador <- sum((tamaño_muestra - 1) * varianza_muestra)  # Calcular el numerador
  suma_tamanos_muestra <- sum(tamaño_muestra, na.rm = TRUE)  # Calcular la sumatoria de
  denominador <- suma_tamanos_muestra - length(tamaño_muestra)  # Calcular el denominado
  if (denominador == 0) {
    return(NaN)  # Si denominador es cero, el resultado es NaN
  } else {
    return(sqrt(numerador / denominador))  # Calcular el resultado final
  }
}

# Calcular LC_Sbarra
LC_Sbarra <- calcular_LC_Sbarra(nuevo_database$Tamaño_Muestra, nuevo_database$Varianza_M

# Imprimir el resultado final
print(paste("Resultado LC_Sbarra:", LC_Sbarra))

#--------------------------------------------------------------------------------

# Calcular la media ponderada de todo el conjunto de datos
LC_Xbarra <- calcular_media_ponderada(Datos6)
A_3 = c(2.659,1.954, 1.628, 1.427)

# Calcular UCL y LCL
UCL <- LC_Xbarra + A_3[nuevo_database$Tamaño_Muestra - 1] * LC_Sbarra
LCL <- LC_Xbarra - A_3[nuevo_database$Tamaño_Muestra - 1] * LC_Sbarra

# Crear un nuevo dataframe con los datos necesarios para graficar
grafica_data <- data.frame(
  Numero_Muestra = 1:34,
  Media_Muestra = nuevo_database$Media_Muestra,
  UCL = UCL,
  LCL = LCL
)

# Graficar la carta de control utilizando ggplot
ggplot(grafica_data, aes(x = Numero_Muestra, y = Media_Muestra)) +
  geom_point(color = 'black', size = 2.5) +
  geom_line(color = '#03396c') +
  geom_line(aes(y = UCL), linetype = 'dashed', color = '#011f4b', lwd = 0.7) +
  geom_line(aes(y = LCL), linetype = 'dashed', color = '#011f4b', lwd = 0.7) +
  ggtitle('Carta para X_barra con muestra variable') +
  ylab('Media muestral') +
  xlab('Número de muestra') +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

```
# Calcular UCL para la carta R
UCL_R <- B_4[nuevo_database$Tamaño_Muestra - 1] * LC_Sbarra

# Crear un nuevo dataframe con los datos necesarios para graficar la carta R
grafica_data_R <- data.frame(
  Numero_Muestra = 1:34,
  Desviacion_Estandar_Muestra = sqrt(nuevo_database$Varianza_Muestra),
  UCL_R = UCL_R
)

# Graficar la carta R utilizando ggplot
ggplot(grafica_data_R, aes(x = Numero_Muestra, y = Desviacion_Estandar_Muestra)) +
  geom_point(color = 'black', size = 2.5) +
  geom_line(color = '#03396c') +
  geom_line(aes(y = UCL_R), linetype = 'dashed', color = '#011f4b', lwd = 0.7) +
  ggtitle('Carta para R con muestra variable') +
  ylab('Desviación estándar muestral') +
  xlab('Número de muestra') +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#################################################################################
################################ CODIGO PUNTO 2 #################################

# Cargar datos
Datos7 <- readxl::read_excel("Datos7.xlsx")
nuevos_nombres <- c("MUESTRA","x1","x2","x3","x4","x5")
names(Datos7) <- nuevos_nombres

### PUNTO 1 ----

# Calcular medias y desviaciones estándar
AA1 <- sqldf::sqldf("SELECT MUESTRA, (x1 + x2 + x3 + x4 + x5) / 5 AS Medias,
                            sqrt(avg(POWER(x1 - (x1 + x2 + x3 + x4 + x5) / 5, 2) +
                                     POWER(x2 - (x1 + x2 + x3 + x4 + x5) / 5, 2) +
                                     POWER(x3 - (x1 + x2 + x3 + x4 + x5) / 5, 2) +
                                     POWER(x4 - (x1 + x2 + x3 + x4 + x5) / 5, 2) +
                                     POWER(x5 - (x1 + x2 + x3 + x4 + x5) / 5, 2))) AS De
                     FROM Datos7
                     GROUP BY MUESTRA")


# Histograma
ggplot(data = AA1, aes(x = Medias)) +
  geom_histogram(binwidth = 1, fill = "#03396c", color = "#011f4b",
                 aes(y = ..density..), alpha = 0.7) +
  geom_density(color = "#005b96") +
  labs(title = "Histograma de Medias", x = "Medias", y = "Densidad") +
  theme_minimal()
```

```
# Q-Q Plot
qqnorm(AA1$Medias)
qqline(AA1$Medias, distribution = qnorm, probs = c(0.25, 0.75), col = "red")
title("Normal Q-Q Plot")

pearson.test(B)

# Gráfico de Control para la Dispersion
Rbar <- mean(AA1$Desviaciones)
D3 <- 0
D4 <- 2.115
UCL <- D4 * Rbar
LCL <- D3 * Rbar

ggplot(data = AA1, aes(x = MUESTRA, y = Desviaciones)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = c(UCL, LCL, Rbar), linetype = "dashed", color = "red") +
  labs(title = "Carta de Control para la Dispersion", x = "Secuencia de Muestreo", y = '
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

# Gráfico de Control para la Media
Xbar <- mean(AA1$Medias)
A2 <- 0.577
UCL <- Xbar + A2 * Rbar
LCL <- Xbar - A2 * Rbar

# Calcular si los puntos están fuera de control
AA1$Fuera_de_Control <- ifelse(AA1$Medias > UCL, 1, 0)
AA1$Fuera_de_Control <- cummax(AA1$Fuera_de_Control)

# Gráfico de Control para la Media
ggplot(data = AA1, aes(x = MUESTRA, y = Medias)) +
  geom_point(aes(color = factor(Fuera_de_Control)), size = 2.5) +
  geom_line() +
  geom_hline(yintercept = c(UCL, LCL, Xbar), linetype = "dashed", color = "red") +
  labs(title = "Carta de Control para la Media", x = "Secuencia de Muestreo", y = "Media
  scale_color_manual(name = "Estado", values = c("black", "red"), labels = c("En control
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

#### PUNTO 2 ------------

# Identificar el punto fuera de control
punto_fuera_de_control <- AA1$MUESTRA[which(AA1$Fuera_de_Control == 1)[1]]

# Calcular la magnitud del cambio
media_antes <- mean(AA1$Medias[1:(punto_fuera_de_control - 1)])
media_despues <- mean(AA1$Medias[punto_fuera_de_control:nrow(AA1)])
```

```
magnitud_cambio <- abs(media_despues - media_antes)

# Imprimir resultados
cat("El proceso se salió de control aproximadamente en la muestra:", punto_fuera_de_cont
cat("La magnitud del cambio fue de:", magnitud_cambio, "\n")

# Graficar las medias con indicación de punto fuera de control y magnitud del cambio
ggplot(data = AA1, aes(x = MUESTRA, y = Medias)) +
  geom_point(aes(color = factor(Fuera_de_Control)), size = 2.5) +
  geom_line() +
  geom_hline(yintercept = c(UCL, LCL, Xbar), linetype = "dashed", color = "red") +
  annotate("point", x = punto_fuera_de_control, y = AA1$Medias[punto_fuera_de_control],
           color = "blue", size = 3) +
  geom_segment(aes(x = punto_fuera_de_control, y = media_antes,
                   xend = punto_fuera_de_control, yend = media_despues),
               color = "blue", linetype = "dotted") +
  annotate("text", x = punto_fuera_de_control + 0.5, y = (media_antes + media_despues) /
           label = paste("Cambio:", round(magnitud_cambio, 2)), color = "blue", size = 3
  labs(title = "Carta de Control para la Media", x = "Secuencia de Muestreo", y = "Media
  scale_color_manual(name = "Estado", values = c("black", "red"), labels = c("En control
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))


### EXTRA -----

# Eliminar el punto fuera de control de la muestra
AA1_sin_punto <- AA1[-which(AA1$Fuera_de_Control == 1)[1], ]

# Calcular nuevamente la media y los límites de control
Xbar_sin_punto <- mean(AA1_sin_punto$Medias)
A2_sin_punto <- 0.577
Rbar_sin_punto <- mean(AA1_sin_punto$Desviaciones)
UCL_sin_punto <- Xbar_sin_punto + A2_sin_punto * Rbar_sin_punto
LCL_sin_punto <- Xbar_sin_punto - A2_sin_punto * Rbar_sin_punto

# Graficar la nueva carta de control para la media
ggplot(data = AA1_sin_punto, aes(x = MUESTRA, y = Medias)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = c(UCL_sin_punto, LCL_sin_punto, Xbar_sin_punto), linetype = "d
  labs(title = "Carta de Control para la Media (sin punto fuera de control)", x = "Secue
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```