

Parcial 2

KEVIN GARCÍA^{1,a}, ALEJANDRO VARGAS^{1,b}

¹DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD DEL VALLE, CALI, COLOMBIA

1. Punto 1

En una planta de potabilización de agua se desea controlar el contenido de plomo (partes por millón) en agua, para ello se ha pensado en construir un gráfico de control, con la estrategia de revisar diariamente a través de la toma de una muestra de 5 unidades. Los resultados se muestran a continuación:

Muestras					
Día	1	2	3	4	5
1	13	8	2	5	8
2	0	6	1	9	15
3	4	2	4	3	4
4	3	15	8	3	5
5	5	10	5	4	0
6	9	5	13	7	7
7	0	4	4	3	9
8	9	3	0	6	0
9	14	0	0	5	3
10	3	9	5	0	2

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

Muestras					
Día	1	2	3	4	5
21	3	7	5	10	12
22	3	0	10	5	4
23	3	3	0	6	9
24	0	2	3	6	7
25	2	3	5	4	10
26	3	1	4	2	4
27	2	4	5	13	4
28	0	16	7	2	11
29	3	5	9	8	6
30	9	7	10	13	0

FIGURE 1: Datos del problema

a. Construir un gráfico de control \bar{X} -barra; R con niveles de significancia del 5% y del 0.27%.

Los dos gráficos de control iniciales $\bar{x} - R$, para los niveles de significancia dados son:

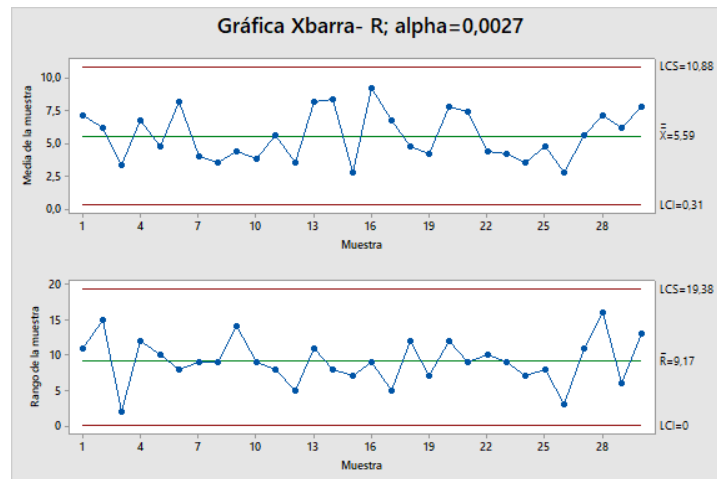
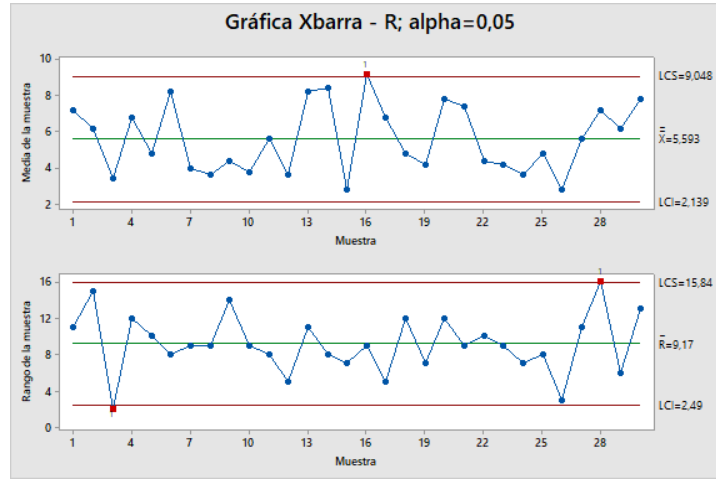


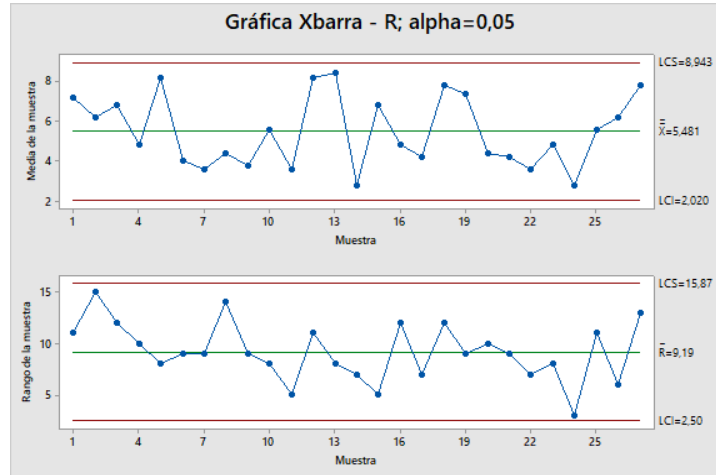
FIGURE 2: Gráfico de control inicial $\bar{x} - R$ para $\alpha = 0.0027$

^aUniversidad del Valle. E-mail: kevin.chica@correounivalle.edu.co

^bUniversidad del Valle. E-mail: jose.alejandro.vargas@correounivalle.edu.co

FIGURE 3: Gráfico de control inicial $\bar{x} - R$ para $\alpha = 0.05$

Podemos observar que en el gráfico $\bar{x} - R$ para $\alpha = 0.05$, tenemos tres puntos por fuera de los límites control, la muestra 16 en el de la media y las muestras 3 y 28 en el del rango, estos se deben eliminar para establecer el gráfico. El gráfico de control luego de eliminar estos dos puntos es:

FIGURE 4: Gráfico de control depurado $\bar{x} - R$ para $\alpha = 0.05$

Se puede notar que el gráfico con un nivel de significancia $\alpha = 0.05$, genera límites más estrictos (estrechos) que su comparativo con $\alpha = 0.0027$.

- b. Simule 1000 subgrupos bajo control, desde la distribución normal y evalúe la significancia de forma empírica para cada gráfico.

A partir de los gráficos de control del literal a, se extrayeron los parámetros que se consideran corresponden a una distribución normal “bajo control”, estos son:

α	X-barra	\bar{R}
0.05	5.48	9.19
0.0027	5.59	9.17

TABLE 1: Tabla para medias.

Con estos parámetros se simularon 1000 subgrupos, una simulación para cada gráfica de control y posterior a esto se procedió a evaluar de forma empírica la significancia; esto se logró contando el número de simulaciones que se salieron de los límites de control LCI y LCS.

Los resultados que se obtuvieron se presentan en la siguiente tabla:

α	X-barra	R
0.05	3.94	3.94
0.0027	3.94	3.94

TABLE 2: Tabla para desviaciones.

	X-barra	R
$\alpha - real$	0.05	0.05
$\alpha - empirico$	0.048	0.046
$\alpha - real$	0.0027	0.0027
$\alpha - empirico$	0.001	0.003

TABLE 3: Tabla comparativa .

Los comentarios se realizarán en el literal d.

- c. Simule 1000 subgrupos con un descentramiento de 2.5 desviaciones estándar, desde la distribución normal y evalúe la potencia de forma empírica para cada gráfico.

Para lograr simular datos con un descentramiento nos apoyamos en la magnitud del cambio, medido en unidades de desviación estándar, el cual viene dado por la formula:

$$r = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Despejando μ_1 (Descentramiento positivo) tenemos:

$$\mu_1 = (2.5 * \sigma_0) + \mu_0$$

Los promedios para cada gráfico están dados por:

α	X-barra
0.05	15.35
0.0027	15.44

TABLE 4: Tabla medias (descentramiento).

Una vez simuladas las poblaciones, se cuenta el número de valores simulados que cayeron dentro de los límites de control, esto significa, cuántas veces el gráfico no detectó el cambio dado que tenemos un proceso fuera de control; posteriormente calculamos la potencia que viene dada por $1 - \beta$ esta nos indica la capacidad del gráfico para detectar correctamente procesos descontrolados.

α	X-barra
0.05	0.999
0.0027	0.991

TABLE 5: Tabla potencias empíricas.

- d. Comente los resultados hallados en (b) y (c), contrastando con las significancias planeadas en (a) y con la potencias teóricas esperadas.

Observamos que en el caso de las significancias para el promedio, se obtuvieron empíricas muy similares, estas dieron como resultado 0.048 y 0.001 para una teórica de 0.05 y 0.0027 respectivamente, lo que quiere decir que de 1000 simulaciones 48 para una significancia de 0.0027 y 1 para la significancia de 0.05 se salieron de los límites de control. En el caso del rango tenemos que las significancias empíricas dieron muy similares a las teóricas, siendo estas 0.046 y 0.003.

Al realizar el cálculo de la potencia de forma empírica, nos dimos cuenta que pese a tener una significancia de 0.0027, la potencia para \bar{x} -barra no deja de ser alta, por lo que podemos decir que el gráfico sigue siendo potente aun cuando tenemos significancia muy baja, esto significa que a la larga se va a tener errores mínimos en las hipótesis, por lo que podemos estar casi seguros que muy rara vez el gráfico se va a equivocar, diagnosticando un proceso fuera de control como bueno y así mismo, muy pocas veces se equivocará diciendo que el proceso está descontrolado cuando la realidad es que todo está en orden, adicional a esto, las potencias empíricas dieron como resultado 0.9952036 para un $K = 3$ y 0.9998584 para un $K = 1.96$.

- e. Con los datos suministrados en la tabla, determine que tan capaz es esta planta de suministrar agua con el estándar mínimo de calidad, si se sabe que para que una fuente de agua pueda ser potabilizada su contenido de plomo no debe ser superior a 10 ppm. Haga su evaluación en el corto y en el largo plazo (recuerde verificar los supuestos básicos del análisis de capacidad).

Primero verificamos si se cumple el supuesto de normalidad, para esto realizamos la prueba Shapiro Wilk, la cual arrojó un p-valor de 0.0000326, adicionalmente, observamos el histograma de frecuencias, donde claramente se aprecia que los datos no siguen una distribución normal.

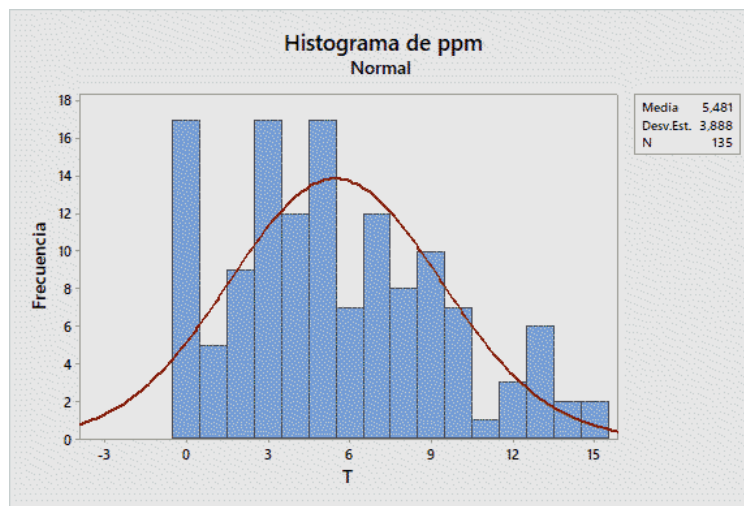


FIGURE 5: Histograma de ppm

Como mis datos no se distribuyen normalmente hay que optar por un cálculo del índice de capacidad para datos no normales, por ende procedemos a evaluar qué tipo de distribución es la que mejor se ajusta a mi conjunto de datos, se obtuvieron algunas de las mejores propuestas.

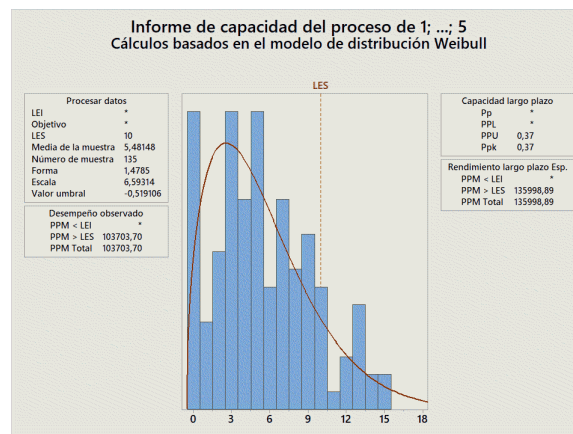


FIGURE 6: Índice de capacidad con Weibul

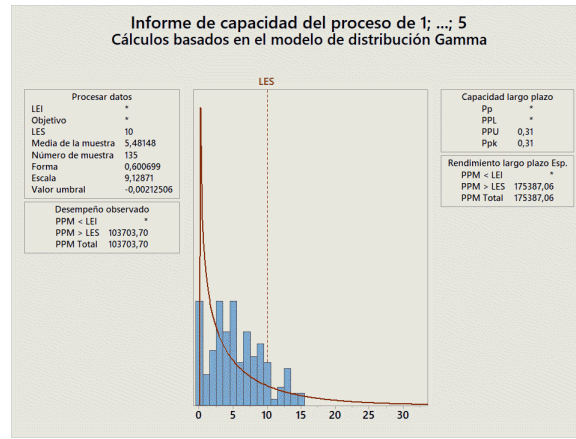


FIGURE 7: Índice de capacidad con Gamma

Estas dos distribuciones son las que “mejor” se acomodan al conjunto de datos, sin embargo vemos que el P_{pk} dio menor que 1 (0.37 y 0.31 respectivamente), lo que indica que el proceso no es capaz, esto quiere decir que la planta con este proceso no es capaz de suministrar agua con el estándar mínimo de calidad.

Observando el rendimiento a corto y largo, en las gráficas se observa que en proporción, aproximadamente el 0.14 de mediciones de ppm en el largo plazo se saldrán de este limite ($LES = 10$), mientras que para el corto plazo, ésta proporción es de aproximadamente el 0.11, estos porcentajes son muy altos ya que de 100 veces, en 14 de ellas el proceso no será capaz de potabilizar el agua.

2. Punto 2

Para controlar el volumen de llenado de un proceso de envasado se ha construido un gráfico de control para el centramiento, obteniendo como resultado $LSC = 1015$; $LIC = 995$. Este gráfico de control ha sido construido con un probabilidad de error tipo I equivalente a 0.6%. Este proceso debe cumplir con un volumen de llenado nominal de 1000 cc con tolerancia de ± 30 ml. Haciendo uso de las observaciones con las que este gráfico de control fue implementado, el controlador de calidad de dicho proceso ha realizado el correspondiente análisis de capacidad, reportando un índice $Cpk = 0,9348$.

- Calcular los límites de control para el gráfico S. (mantener el mismo nivel de error tipo I y el mismo tamaño de muestra)

La información dada es:

$$LIC = 995 \text{ (Limite de control inferior)}$$

$$LSC = 1015 \text{ (Limite de control superior)}$$

$$LIE = 970 \text{ (Limite inferior especificado)}$$

$$LSE = 1030 \text{ (Limite superior especificado)}$$

$$\mu = 1000$$

$$\alpha = 0.006$$

$$Cpk = 0.9348$$

A partir de esto, debemos hallar n y \bar{S} para poder construir el gráfico S.

Tenemos el siguiente sistema de ecuaciones, teniendo en cuenta que $\alpha = 0.006$, entonces $k = 2.75$:

$$995 = \bar{\bar{X}} + 2.75 \frac{\sigma}{\sqrt{n}} \quad (1)$$

$$1015 = \bar{\bar{X}} - 2.75 \frac{\sigma}{\sqrt{n}} \quad (2)$$

$$LC = \bar{\bar{X}}$$

Sumando (1) y (2) tenemos:

$$\begin{aligned} 2010 &= 2\bar{\bar{X}} \\ \bar{\bar{X}} &= 1005 \end{aligned}$$

También sabemos que

$$C_{pk} = \text{Min}(C_{pu}, C_{pl})$$

Donde:

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

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

En este caso $C_{pk} = C_{pu}$, ya que analizando los numeradores $LSE - \mu = 25 < \mu - LIE = 35$, por tanto $\text{Min}(C_{pu}, C_{pl}) = C_{pu}$

Como ya conocemos el valor de C_{pk} :

$$0.9348 = \frac{1030 - 1005}{3\sigma} \rightarrow 2.8044\sigma = 25 \rightarrow \hat{\sigma} = \frac{25}{2.8044} = 8.9145$$

Reemplazando $\hat{\sigma} = 8.9145$ en (1), obtenemos n:

$$1015 = 1005 + 2.75 \frac{8.9145}{\sqrt{n}} \rightarrow \sqrt{n} = 2.4515 \rightarrow n = 6$$

Dado que según Montgomery, se puede trabajar $\hat{\sigma} = \bar{S}/c_4$, ya que la varianza del proceso es desconocida. Y ya sabemos que $n = 6$, entonces c_4 de acuerdo a las tablas tabuladas es igual a 0.9515, entonces, reemplazando:

$$8.9145 = \frac{\bar{S}}{0.9515} \rightarrow \bar{S} = 8.4821$$

Finalmente, los limites para el gráfico S son:

$$LCS = \bar{S} + 2.75 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} = 16.024$$

$$LC = \bar{S} = 8.4821$$

$$LCI = \bar{S} - 2.75 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} = 0.94$$

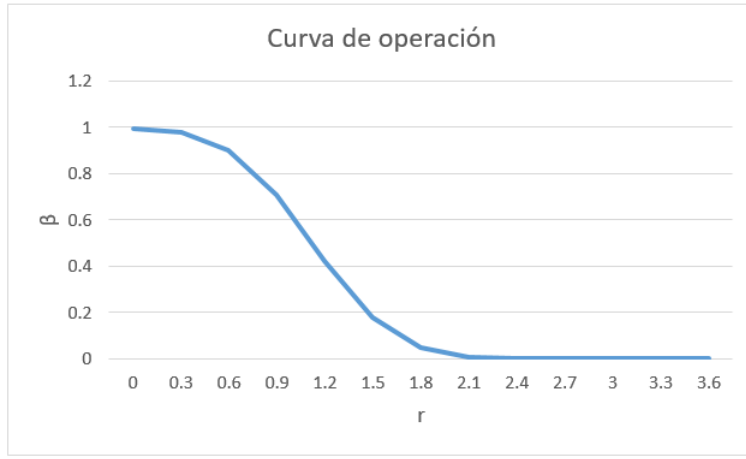
Dado que el n y el k en ambos casos es de 6 y de 2.75 respectivamente, se asegura que en el gráfico S se mantiene el mismo nivel de error tipo I y el mismo tamaño de muestra

- b. Construir y comentar la curva característica de operación para el gráfico X barra asociado y su curva ARL. (evaluar en el rango 0 - 3.5 desviaciones estándar)

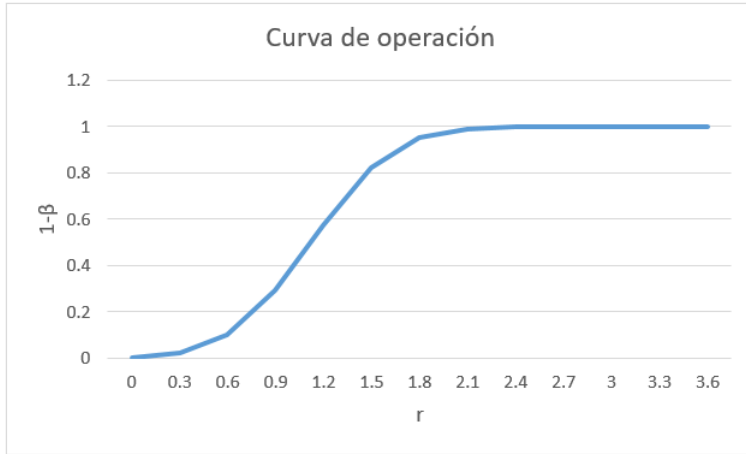
Para construir la curva de operación, se calcularon las probabilidades de no detectar un cambio dado que el proceso está fuera de control, para diferentes magnitudes de cambio r de 0 a 3.5%. Estas probabilidades se calcularon mediante:

$$\beta = P(\text{No detectar un cambio} | \text{Fuera de control})$$

$$\beta_r = P(Z < k - r\sqrt{n}) - P(Z < -k - r\sqrt{n})$$

FIGURE 8: Curva característica de operación(β_r)

Para realizar más fácil la interpretación, se obtuvo y se graficó $1 - \beta_r$ que es la probabilidad de detectar el cambio, dado que el proceso está fuera de control.

FIGURE 9: Curva característica de operación($1 - \beta_r$)

Se puede apreciar que para cambios de una magnitud r baja, entre 0 y 0.7%, la probabilidad de detectar cambios de este tipo está entre 0 y 0.2 de forma ascendente. Como es de esperarse, mientras más grande es la magnitud del cambio r , más alta es la probabilidad de que el gráfico de control lo detecte, por ejemplo, a partir de cambios de una magnitud r de 0.7% hasta 1.5%, la probabilidad de detectar el cambio crece casi que linealmente. Finalmente, estas probabilidades se estabilizan en 1, cuando la magnitud del cambio es mayor o igual a 2, esto quiere decir que cambios de estas magnitudes, siempre los detectará el gráfico de control.

Para construir la curva ARL, se calculará para r entre 0 y 3.5:

$$ARL_r = \frac{1}{p_r} = \begin{cases} \frac{1}{\alpha} & r = 0 \\ \frac{1}{1 - \beta_r} & r > 0 \end{cases}$$

Con respecto al ARL (número promedio de observaciones que dejará pasar el gráfico antes de emitir una señal bajo un cambio de magnitud $r\sigma$, se observa en el gráfico, que cuando el proceso está bajo control, es decir, $r = 0$, la frecuencia promedio de aparición de falsas alarmas es de aproximadamente 160 muestreos, mientras que cuando el proceso está fuera de control a una magnitud r baja (entre 0.1 y 0.6%), la frecuencia promedio de detección del cambio va desde 140 muestreos en el mínimo ($r = 0.1$) hasta 10 muestreos en el máximo ($r = 0.6$), a partir de una magnitud de cambio de 0.9% en adelante, la frecuencia promedio de detección del cambio se estabiliza en 0, es decir, que el cambio siempre se va a detectar.

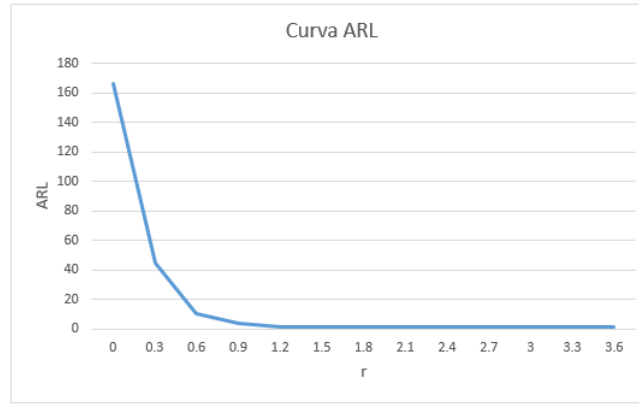


FIGURE 10: Curva ARL

- c. En determinado momento el proceso de llenado comienza a envasar botellas con un promedio de 984 cc. Evalúe la probabilidad de que el gráfico detecte este cambio antes de 3 inspecciones. Para calcular la probabilidad de que el gráfico detecte un cambio en la media del proceso de una magnitud r , debemos calcular el complemento de:

$$\beta = P(\text{No detectar un cambio} | \text{Fuera de control})$$

$$\beta_r = P(Z < k - r\sqrt{n}) - P(Z < -k - r\sqrt{n})$$

Debemos calcular la magnitud del cambio r que corresponde al cambio hasta 984 cc

$$r = \frac{|\mu_1 - \mu_0|}{\sigma_0} = \frac{|984 - 1005|}{8.9145} = 2.35$$

Obtenida la magnitud del cambio $r = 2.35\%$, y tenemos que para $\alpha = 0.006$, $k = 2.75$. Por lo tanto:

$$\beta_{2.35} = P(Z < 2.75 - 2.35 \cdot \sqrt{6}) - P(Z < -2.75 - 2.35 \cdot \sqrt{6}) = P(Z < -3) - P(Z < -8.5) = 0.00135 - 0 = 0.00135$$

Entonces, la probabilidad de detectar el cambio será:

$$1 - \beta_{2.35} = 1 - 0.00135 = 0.99865$$

Finalmente, la probabilidad de que el gráfico detecte el cambio antes de 3 inspecciones (en 2 o menos) es, la probabilidad de que lo detecte en la primera más la probabilidad de que lo detecte en la segunda:

$$(1 - \beta_{2.35}) + \beta_{2.35}(1 - \beta_{2.35}) = 0.99865 + (0.00135)(0.99865) = 0.999998$$

Se puede decir que como la magnitud del cambio es considerablemente grande, el gráfico lo detectará casi siempre (99.86% de las veces) en la primera inspección, y casi el 100% de las veces lo detectará antes de 3 inspecciones.

- d. Si se presenta un cambio como el enunciado en c, cuál sería el porcentaje de unidades no conformes que se producirían?, cuál sería el nuevo valor Cpk y Cp.

Si se presenta un cambio como el enunciado en c, es decir, si μ pasa a ser 984 cc, el porcentaje de unidades no conformes que se producirían la podemos encontrar como el complemento de:

$$\begin{aligned} P(970 < \bar{x} < 1030 | \mu = 984) &= P(\bar{x} < 1030 | \mu = 984) - P(\bar{x} < 970 | \mu = 984) \\ &= P\left(Z < \frac{1030 - 984}{8.9145/\sqrt{6}}\right) - P\left(Z < \frac{970 - 984}{8.9145/\sqrt{6}}\right) \end{aligned}$$

$$= P(Z < 12.64) - P(Z < -3.85) = 1 - 0.00006 = 0.99994$$

Entonces, se producirán $1 - 0.99994 = 0.00006$ unidades no conformes, es decir, se producirán 60 unidades no conformes por millón.

El valor de Cp no cambia, ya que no depende de la media del proceso, este valor es:

$$Cp = \frac{LSE - LIE}{6\sigma} = \frac{1030 - 970}{6(8.9145)} = 1.1217$$

Lo que nos estaría diciendo que el proceso es capaz con necesidad de vigilancia.

El nuevo valor Cpk es:

$$C_{pk} = \min(C_{pu}, C_{pl})$$

Donde:

$$C_{pu} = \frac{LSE - \mu}{3\sigma} = \frac{1030 - 984}{3(8.9145)} = 1.72$$

$$C_{pl} = \frac{\mu - LIE}{3\sigma} = \frac{984 - 970}{3(8.9145)} = 0.52$$

Por tanto:

$$C_{pk} = \min(1.72, 0.52) = 0.52$$

Lo que nos indica que el proceso no cumple con por lo menos una de las especificaciones.