

LA RESIGNIFICACIÓN DEL CONCEPTO DE PROBABILIDAD Y LA TOMA DE DECISIONES. LOS CONTROLES DE CALIDAD.

Mag. Alicia Pereira das Neves Yedig
apereiradasneves@gmail.com
Ce.R.P. del Litoral - Uruguay

Tema: Pensamiento relacionado con la Probabilidad y la Estadística

Modalidad: Conferencia regular

Nivel educativo: Medio (11 a 17 años)

Palabras clave: Probabilidad, frecuencias relativas, variabilidad y control de calidad.

Resumen

Lograr la mejora de la calidad en los procesos de producción industrial, ha requerido analizar las causas de la variabilidad que en ellos se genera.

Los controles de calidad en muchos casos, se basan en diseños de muestras pequeñas. Son ejemplos para abordar en el aula los conceptos de Estadística Descriptiva desde Ciclo Básico y/o avanzar hacia los de Estadística Inferencial en Bachillerato. A partir de los contrastes de normalidad realizados y de los test de rachas, se resignifica probabilidad. Los gráficos de control requieren del uso de software disponible y al alcance de nuestros alumnos (Excel, R, etc.).

La variabilidad en los procesos y la Mejora en la Calidad:

La identificación de los problemas de Calidad en las etapas de producción y distribución de los productos manufacturados ocupa el centro de atención de los ingenieros y personal especializado.

Elevar los niveles de Calidad, se sabe que disminuye los costos e incrementa la satisfacción de los clientes.

Soluciones Estadísticas:

Los gráficos de control y métodos de muestreo de lotes para aceptación, desarrollados en los años 1920, se utilizan en la industria como diagnóstico para identificar en forma muy sencilla, circunstancias inusuales. Además de esto, recientemente el japonés G. Taguchi y sus seguidores han desarrollado métodos estadísticos basados en fracciones factoriales, que han implicado una gran innovación en la línea de investigación estadística sobre estos temas.

Causas comunes y causas asignables:

Es claro que reducir la variabilidad en el proceso es equivalente a mejorar la calidad del proceso, sin embargo, más allá del control y calibración de las máquinas, de los factores ambientales y materiales, de los operarios, etc, se producirá variación natural inherente al proceso, que no es controlable.

Otro tipo de variación de impacto negativo en la calidad del proceso, es atribuible por ejemplo a material contaminado o mala calidad de la materia prima, ubicación incorrecta de los controles de las máquinas, mal funcionamiento de una máquina, indiferencia de un operario, etc.

W. A. Shewart clasificó las causas de las variaciones en:

Causas comunes: Inherentes al proceso, difíciles de eliminar, cúmulo de muchas pequeñas variaciones pero que sin embargo representan más del 85% de los problemas de Calidad.

Causas especiales o asignables: No son parte del proceso, son detectables con herramientas estadísticas como los gráficos de control, fáciles de eliminar por los operarios en el proceso y, aunque suelen representar menos del 15% de los problemas de Calidad, producen efectos importantes a largo plazo.

Cuando el proceso sólo está afectado por causas comunes, se dirá que el proceso está en estado de control estadístico, o bajo control.

Gráficos de Control:

Los gráficos de control son la principal técnica de control estadístico de procesos y proporcionan un mecanismo para reconocer situaciones donde las *causas asignables* podrían afectar de manera adversa la calidad del producto. Detectada una situación fuera de control, puede profundizarse la investigación para reconocer las causas y ajustar o corregir el proceso.

Si es preciso estudiar por ejemplo en un proceso el valor de una variable continua, usualmente se seleccionan en forma periódica, muestras aleatorias simples pequeñas y de tamaño fijo n del proceso de interés, con lo que se obtiene para cada una un valor de un estadístico de interés como puede ser la media, rango y varianza muestral. Así, será posible realizar un gráfico sencillo, de los valores medios de las muestras en relación al

número de la muestra o tiempo en que se seleccionó. En cualquier caso, surge como importante tener en cuenta la variabilidad de los datos dentro de la pequeña muestra, que queda muy bien descrita por el rango, dado que se trata de muestras de pequeño tamaño. Es posible jerarquizar desde estos ejemplos sencillos, la variabilidad de los datos y su importancia en la descripción de los mismos, como contenido explícito en el aula.

Dado que la inferencia se basa en cálculos probabilísticos y se tomarán decisiones en contextos de incertidumbre, es importante tener en cuenta que es posible que un proceso se juzgue fuera de control cuando se encuentra bajo control o, afirmar que está bajo control siendo que en verdad no lo está. Este tipo de errores deben ser tenidos en cuenta y analizados siempre que de inferencias estadísticas se trate. Abordarlos como contenidos desde los primeros desarrollos conceptuales en Estadística promoverá otros niveles de comprensión de Estadística Inferencial a través de actividades adecuadas. Las limitaciones de las inferencias estadísticas se reducen notablemente si se tiene en cuenta que la disciplina se constituye en ciencia porque es capaz de medir en términos de probabilidad, los errores que se cometen al realizarlas.

Marco conceptual:

Supóngase que la variable estudiada X , que puede hacer referencia al diámetro de los tornillos fabricados en una planta de producción, resistencia de un cable eléctrico, etc., tiene una distribución normal con media μ y varianza σ^2 , entonces, si \bar{X} es la media muestral para una de las muestras de tamaño n , seleccionada en determinado instante de tiempo, se tiene en virtud del Teorema Central del Límite que:

\bar{X} tiene una distribución normal.

$$E(\bar{X}) = \mu$$

$$\text{Var}(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \quad \text{y por lo tanto:}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

De esta forma la variable que hace referencia al valor medio muestral \bar{X} puede ser estandarizada de la siguiente forma: $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1)$

Si se ha abordado la distribución normal para una variable a partir del concepto de probabilidad y el de esta como una medida de ocurrencia de un suceso, se podrá analizar cuál o cuáles son algunos de los sucesos que se espera ocurran con baja frecuencia y cuáles con alta frecuencia.

Así por ejemplo, puede observarse a partir de cálculos sencillos de probabilidad, que hay ciertos sucesos que, bajo el supuesto de normalidad de la variable en estudio y con ciertos parámetros poblacionales, tendrán baja o casi nula probabilidad de ocurrir y otros serán casi seguros. Será oportuno analizar las frecuencias relativas con que estos ocurren en la realidad, efectuando contrastes con los cálculos teóricos para así tomar decisiones acerca de los supuestos de hipótesis. No obstante este análisis requiere una validación posterior de la normalidad de la variable que se ha utilizado como modelo de referencia.

Conociendo la distribución de la variable Z , normal estándar, proponemos un suceso casi seguro:

$$p(\mu - 3\sigma_{\bar{X}} \leq \bar{X} \leq \mu + 3\sigma_{\bar{X}}) = p(\mu - 3\sigma_{\bar{X}} - \mu \leq \bar{X} - \mu \leq \mu + 3\sigma_{\bar{X}} - \mu) =$$

$$= p\left(\frac{-3\sigma_{\bar{X}}}{\sigma_{\bar{X}}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{3\sigma_{\bar{X}}}{\sigma_{\bar{X}}}\right) = p(-3 \leq Z \leq 3) \cong 0.9974$$

El cálculo anterior que requiere nociones básicas de la distribución normal que se enseña en Bachillerato, permite comprender por qué es de esperar en un proceso de producción bajo control y en el que las únicas variaciones de los datos son debidas a causas comunes y no especiales o asignables, que un valor medio muestral quede contenido dentro de tres desvíos estándares. Luego se concluyen como límites de control inferior y superior respectivamente a los siguientes:

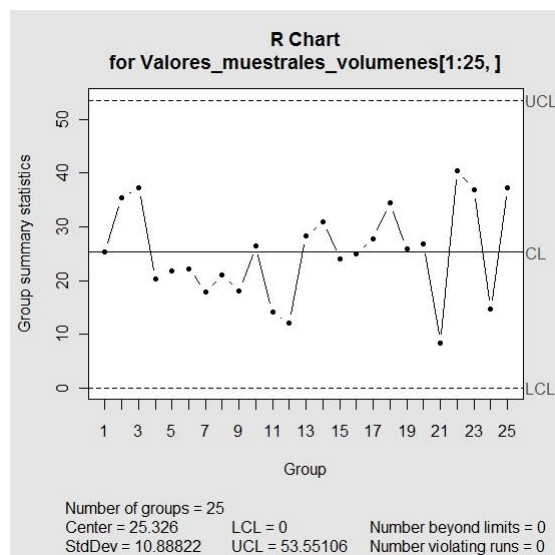
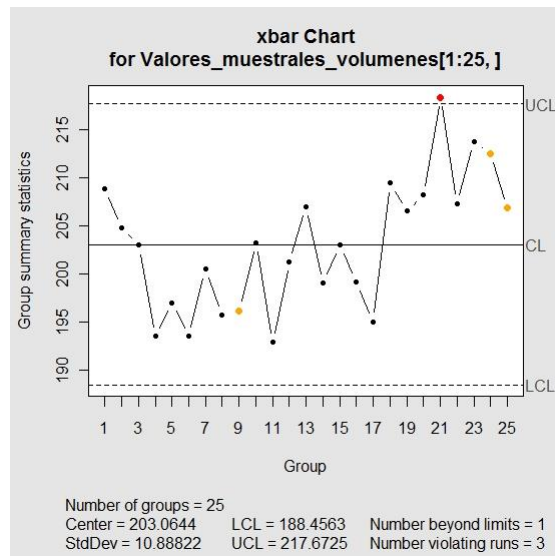
$$LIC = \mu - 3\sigma_{\bar{X}} = \mu - 3\frac{\sigma}{\sqrt{n}} \quad \text{y} \quad LSC = \mu + 3\sigma_{\bar{X}} = \mu + 3\frac{\sigma}{\sqrt{n}}$$

Propuestas para el aula:

En la siguiente tabla se recogen los datos procedentes de un control del proceso de envasado de yogures de 200 ml.:

MUESTRA					
1	201.72	204.52	227.07	206.59	204.09
2	200.06	193.35	224.97	189.49	216.02
3	221.94	215.59	198.38	184.61	194.38
4	194.92	192.28	190.42	205.12	184.85
5	191.09	211.90	197.68	190.03	194.34
6	193.50	187.54	209.27	187.03	190.12
7	203.71	207.26	203.44	198.66	189.41
8	210.19	189.20	190.47	194.58	194.19
9	189.15	198.19	199.68	187.92	206.04
10	200.52	210.57	217.08	190.58	197.57
11	201.34	190.96	187.65	187.25	197.18
12	204.69	206.92	202.58	194.92	197.15
13	219.77	207.20	202.03	214.50	191.40
14	208.13	196.31	212.58	181.60	196.58
15	210.11	204.17	198.88	188.99	213.03
16	201.15	181.60	203.25	203.33	206.65
17	192.42	203.70	205.24	177.40	196.30
18	193.76	204.42	217.45	203.41	228.16
19	214.24	210.65	209.02	210.29	188.43
20	196.42	201.52	210.74	223.21	209.18
21	220.80	213.80	222.07	217.66	217.46
22	227.70	206.43	203.55	187.18	211.70
23	217.54	203.25	199.80	236.78	211.28
24	203.97	217.71	218.62	217.82	204.58
25	193.21	224.67	217.31	211.71	187.31

Se ha realizado un control del proceso, seleccionando 25 muestras de 5 unidades cada una. Se nos ha informado que la distribución para la variable *volumen* de uno de los yogures, puede suponerse normal con media $\mu=200$ y desvío estándar $\sigma=10$. En este caso, con la ayuda del paquete estadístico *R*, se han procesado los datos obtenidos, realizándose los gráficos que se registran en las figuras siguientes:



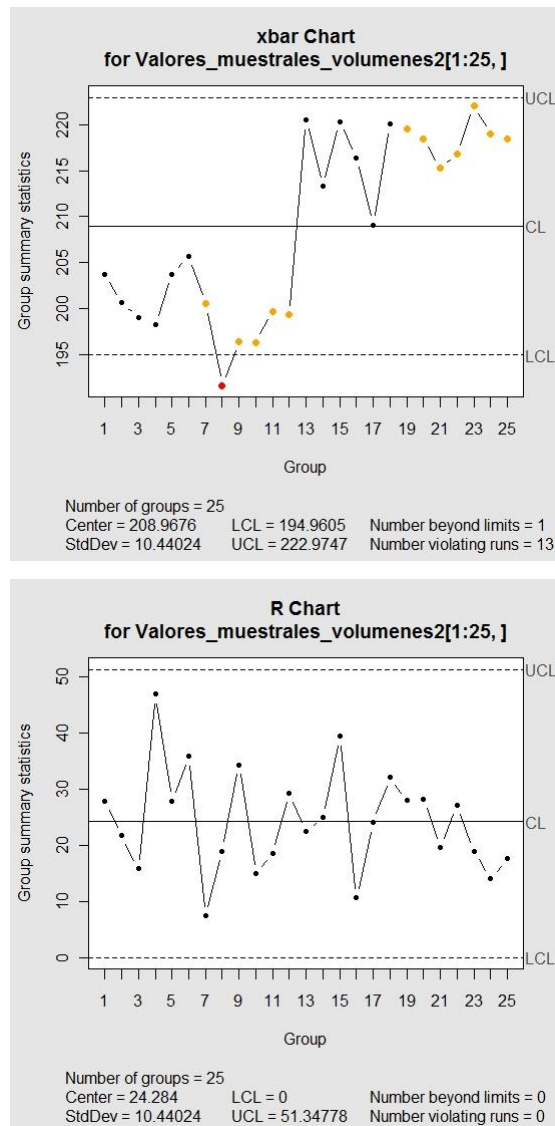
Resulta de interés, analizar detenidamente estos gráficos que pueden hacerse con la ayuda de este u otro software que esté al alcance de los estudiantes, contrastando caso a caso los valores medios y los rangos en cada una de las 25 muestras. Esto arrojará luz al análisis que se está realizando, dado que se tienen parámetros de control para los valores medios, su variabilidad global y variabilidad interna en cada muestra. Téngase en cuenta que cada valor medio muestral se cruzará con el valor correspondiente al rango muestral, que dado el tamaño muestral reducido, brinda mucha información acerca de la dispersión de los valores dentro de cada muestra.

El docente en su curso podrá optar por realizar alguno de estos gráficos en forma manual, para abordar los contenidos procedimentales que considere pertinente, identificando cuál es la variable, clasificándola, realizando un resumen de sus valores, analizando por qué se busca ese tipo de representación gráfica (valores medios según el índice de la muestra), cuáles son las escalas adecuadas, etc. Luego, el uso de las herramientas que este u otro software específico nos aportan, puede optimizar el uso del tiempo áulico en aras de profundizar el análisis e interpretación de resultados.

Si el control del proceso hubiese recogido los siguientes datos, es probable que el análisis no hubiese sido el mismo. Observemos los puntos que se señalan y definamos algunos sucesos para los cuales en cursos superiores es sencillo el cálculo de probabilidades:

MUESTRA

1	184.39	209.80	212.21	204.49	207.59
2	208.35	189.63	200.19	211.39	193.59
3	195.38	207.71	191.90	205.38	194.82
4	183.70	215.46	172.29	200.42	219.28
5	200.65	203.59	215.69	211.03	187.75
6	199.65	228.93	210.09	196.62	193.01
7	204.44	201.88	197.82	201.47	197.01
8	185.44	200.40	200.80	189.59	181.80
9	183.91	218.26	206.51	184.61	188.63
10	191.34	203.85	188.94	194.19	203.12
11	201.65	199.03	191.43	196.48	210.01
12	210.75	196.65	213.64	191.12	184.37
13	213.15	212.56	211.19	232.07	233.64
14	226.30	208.53	216.82	213.68	201.35
15	218.62	219.26	202.58	219.33	241.99
16	220.80	210.56	221.24	215.29	214.12
17	198.86	221.79	216.43	197.71	210.71
18	211.18	214.98	207.03	227.91	239.22
19	210.36	213.61	211.71	223.53	238.32
20	208.02	227.71	227.94	228.50	200.29
21	215.36	211.06	224.01	204.35	221.68
22	221.13	203.36	218.45	210.77	230.49
23	228.06	218.39	227.83	209.16	226.90
24	218.82	215.96	213.80	227.85	218.35
25	212.23	220.01	215.02	215.02	229.88



Análisis de Patrones no Aleatorios:

Los test de rachas o patrones de causas asignables y que en general vienen predefinidos en la mayoría de los paquetes estadísticos, son los que se proponen a continuación.

Estos son sólo algunos casos que tienen probabilidad muy baja de ocurrir bajo el supuesto de normalidad de la variable y aleatoriedad, sin causas especiales que los generen. Si se detecta alguna de estas situaciones se considerará que es razón suficiente para admitir que no es aleatoria y que sí tiene una causa especial que la genera, lo que clasifica el proceso como fuera de control.

Consideremos los siguientes conjuntos, cuya representación gráfica se sugiere realizar, sin demasiada formalización algebraica:

$$A = \{(x, y) \in R^2 : (\mu - 3\sigma_{\bar{x}} \leq y \leq \mu - 2\sigma_{\bar{x}}) \vee (\mu + 2\sigma_{\bar{x}} \leq y \leq \mu + 3\sigma_{\bar{x}})\}$$

$$B = \{(x, y) \in R^2 : (\mu - 2\sigma_{\bar{x}} \leq y \leq \mu - \sigma_{\bar{x}}) \vee (\mu + \sigma_{\bar{x}} \leq y \leq \mu + 2\sigma_{\bar{x}})\}$$

$$C = \{(x, y) \in R^2 : \mu - \sigma_{\bar{x}} \leq y \leq \mu + \sigma_{\bar{x}}\}$$

Se consideran los siguientes sucesos, que hacen referencias a la representación gráfica de los valores medios indexados por su número de muestra:

- Un punto más allá de la zona A.
- Dos de tres puntos consecutivos en la zona A o más allá.
- Cuatro de cinco puntos consecutivos en la misma zona B o más allá.
- Nueve puntos consecutivos en la misma zona C o más allá.
- Quince puntos consecutivos en la zona C (a ambos lados de la línea central), lo que seguramente indica que σ ha sido sobre-dimensionada.
- Ocho puntos seguidos a ambos lados de la línea central, pero ninguno en C.
- Catorce puntos consecutivos alternativamente arriba y abajo.
- Seis puntos seguidos permanentemente crecientes o decrecientes.

¿Con qué probabilidades ocurrirían estos sucesos por azar? ¿Cuáles de estos sucesos han ocurrido en el último diseño muestral?

Referencias bibliográficas

Batanero, C. y Díaz, C. (2010). *Estadística con Proyectos*. Disponible en: <http://www.ugr.es/~batanero/ARTICULOS/Libroproyectos.pdf>

Canavos, G. (1997). *Probabilidad y Estadística*. Madrid: Mc Graw Hill.

Devore, J. (2005). *Probabilidad y Estadística para ingeniería y ciencias*. Mexico: Thomson Editores.

Fernández, F. y Mayor, J. (1995). *Muestreo en poblaciones finitas*. Barcelona: EUB.

Galmés, M. (1997). *Métodos de muestreo*. Montevideo: Oficina de Apuntes del CECEA. Facultad de Ciencias Económicas y Administración.

Pereira das Neves, A. (2010). *Convergencia de variables aleatorias: teoría y aplicaciones*. Proyecto de Licenciatura. Dirigido por la Dra. Ma. Isabel Marrero Rodríguez.