Tarea 14
Teorema del límite central

Erick Cervantes Mendieta
Matrícula: 2032430

Modelos Probabilistas Aplicados

08/12/2020

1. Conceptos

El **teorema del límite central** nos indica que si el tamaño de una muestra es lo *suficientemente grande* ($n > 30$), la distribución de medias de muestra puede aproximarse a una distribución normal, aun si la población original no se distribuye normalmente [1]. El teorema de límite central se puede aplicar a una *muestra aleatoria* X_1, X_2, …, X_n para cualquier distribución mientras $\mathbb{E}(X_i) = \mu$ y $\mathbb{V}(X_i) = \sigma^2$ sean finitas y el tamaño muestral sea grande.

La media de todas las medias de muestra es la media poblacional μ y la desviación estándar de todas las medias de muestra es $\sigma/\sqrt{n}$. El teorema del límite central implica dos distribuciones diferentes: la distribución de la población original y la distribución de las medias de muestra.

Es posible resolver algunos problemas prácticos con el teorema del límite central, y cuando se trabaje con ese tipo de problemas es bueno recordar que si se tiene un valor individual de una población que se distribuye normalmente, se debe utilizar $z = (x - \mu)/\sigma$, y cuando se trabaje con una media de alguna *muestra* (o grupo), utilizar el valor de $\sigma/\sqrt{n}$ para la desviación estándar de las medias de muestra (utilice $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$).

2. Aplicación

El **tratamiento del agua de una caldera** de vapor es fundamental para asegurar una larga vida útil libre de problemas operacionales, reparaciones y como prevención de algunos accidentes que se pudieran ocasionar. Así, el objetivo es evitar problemas de corrosión e incrustaciones, asegurando la calidad del agua de alimentación y del agua contenida en la caldera.

En el Ingenio Central Motzorongo, ubicado en el estado de Veracruz, se maneja un formato para medir algunas variables en el tratamiento interno del agua en las diferentes calderas (caldera 5, 7, 8 y 9), como lo es el pH del agua (pH), la alcalinidad determinada con fenolftaleína (F), la alcalinidad determinada con naranja de metilo (M), hidróxidos (OH), cloruros (CL), fosfatos (PO_4), sulfitos (SUL) y los sólidos totales disueltos (STD).

Estas mediciones tiene el objetivo de verificar que los valores de las variables medidas estén en *control*, es decir, que los valores estén dentro de las especificaciones que marca el ingenio, ya que si no se logra esto, entonces los encargados procederán a dar un *tratamiento* adecuado al agua, el aseguramiento de la calidad del agua de alimentación y agua de la caldera se consigue cumpliendo con los requerimientos de las normas, el ingenio ha estipulado algunos parámetros para dicho tratamiento, como por ejemplo, para el valor de pH, se tiene que los valores leídos puede variar entre valores de 10 a 11.5, ya que si oscilan fuera de estos valores se podrían tener problemas de corrosión (bajo pH) y/o depósitos (alto pH). Bajo este supuesto, se puede indicar que la desviación estándar es igual a 0.75.

Durante la *zafra* pasada se capturaron 100 mediciones (observaciones) para las diferentes calderas, se desea saber si hay evidencia para sugerir que los valores de pH encontrados en la caldera 5 (véase la base de datos `caldera5.txt`) no cumplen con las especificaciones que ha impuesto el ingenio.

Para esto, definimos a $\bar{X}$ como el valor promedio observado al medir el pH en el agua de la caldera, aunque no se sabe si los datos provienen de una distribución normal, pero como n excede a 30, entonces se puede utilizar el teorema de límite central para calcular la probabilidad de que $10 \leq \bar{X} \leq 11.5$, es decir, se desea calcular $P(10 \leq \bar{X} \leq 11.5)$. Luego, se tiene que $\bar{X} = 11.195$. Luego,

$$
\begin{aligned}
P(10 \leq \bar{X} \leq 11.5) &= P\left(\frac{11.195 - 10}{0.75/\sqrt{100}} \leq Z \leq \frac{11.195 - 11.5}{0.75/\sqrt{100}}\right) \\
&= P\left(\frac{1.195}{0.075} \leq Z \leq \frac{-0.305}{0.075}\right) \\
&= P(15.93 \leq Z \leq -4.07) \\
&\approx 1 - P(15.93 \leq Z) + P(Z \leq -4.07) \\
&= 1 - 0.9999 + 0.0001 = 0.0002,
\end{aligned}
$$

debido a que esta probabilidad es muy baja, la evidencia sugiere que los valores obtenidos al medir el pH en la caldera 5 no se encuentran dentro del rango especificado por el ingenio. Es decir, es probable que el tratamiento de agua de la caldera no fue el adecuado o estuvo muy variante durante toda la zafra (ver figura 1).

En la figura 2 se puede observar el comportamiento de 40 mediciones de tamaño 4 de la caldera 7, en donde se puede apreciar como la distribución de las medias va tomando una forma de distribución normal.

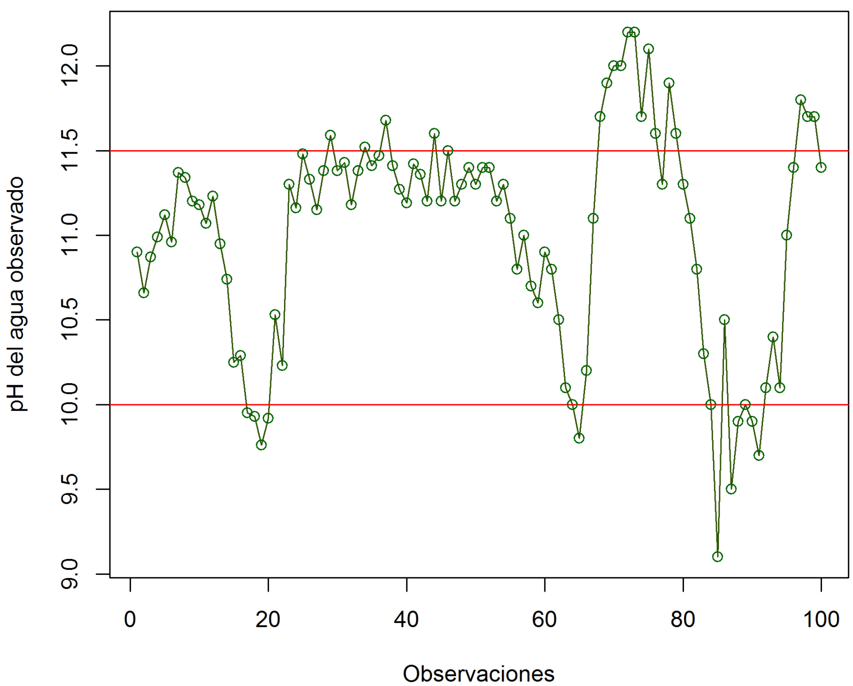

Figura 1: Valores observados del pH del agua en la caldera 5.

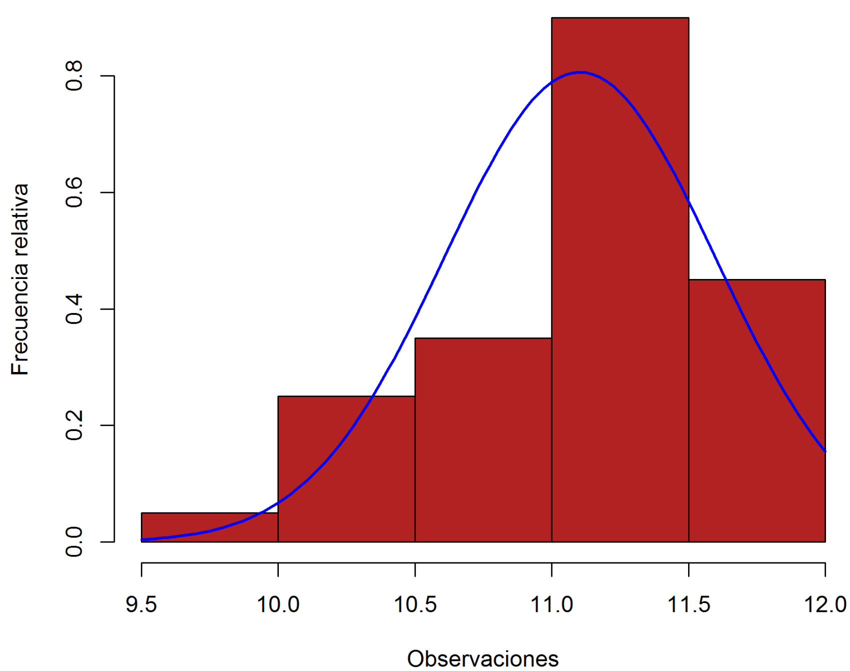

Figura 2: Valores observados del pH del agua en la caldera 7 (40 muestras de tamaño 4).

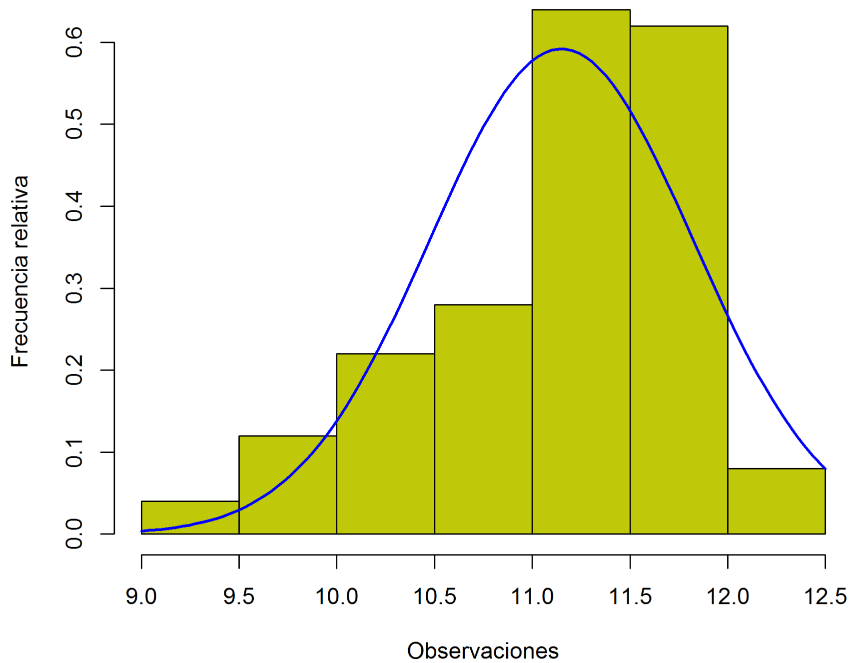

Figura 3: Valores observados del pH del agua en la caldera 5 (100 muestras de tamaño 3).

Por otra parte, en la figura 3, se aprecia como la distribución de las medias se suaviza aún más, conforme se va aumentando el tamaño de las muestras, en este caso se consideraron 100 observaciones de tamaño 3 para la caldera 5, sin embargo, aún no es suficiente ver el comportamiento para visualizar el teorema del límite central, ya que el tamaño de las muestras debe ser demasiado grande.

Referencias

[1] M. F. Triola. *Estadística*. Pearson Education, México, 2004.