

Control 3

KND

21 de octubre de 2020

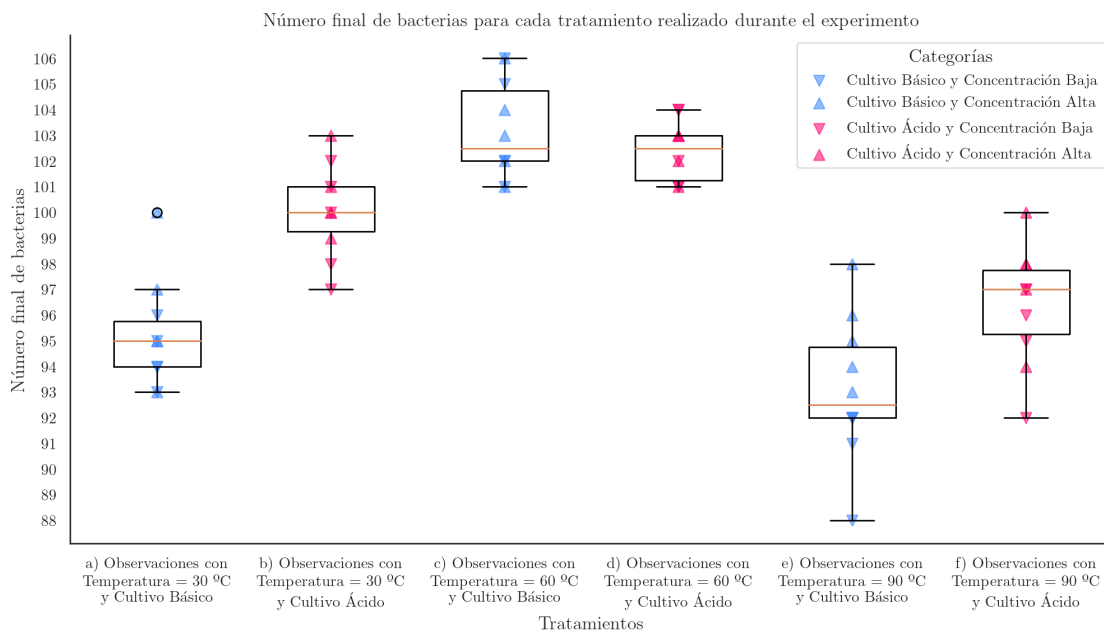
1 Introducción

En un laboratorio se llevó a cabo un experimento bajo distintas condiciones para determinar la influencia que el nivel de temperatura y la naturaleza ácida de cierto caldo de cultivo tienen sobre el crecimiento de una población de bacterias.

En este documento se plantea, se ajusta y se analiza un modelo de regresión lineal a partir de los resultados del experimento tomando como respuesta al número final de bacterias después de 48 horas de una población inicial de 1000 y como regresores a los factores de Temperatura (30 °C, 60 °C o 90 °C), tipo de Cultivo (Básico o Ácido) y Concentración (Baja o Alta) con el propósito de obtener conclusiones sobre los resultados del experimento realizado.

2 Describa los datos gráficamente.

Como los factores de interés son Temperatura (30 °C, 60 °C o 90 °C) y Cultivo (Básico o Ácido), se considera un total de 6 tratamientos distintos, cada uno con dos niveles de Concentración (Baja o Alta). En la Figura 1. se encuentran las gráficas de caja y brazo (*boxplots* en inglés) para cada uno de los tratamientos considerados. En el eje vertical se encuentra el Número final de bacterias, es decir, la respuesta, y en el eje horizontal se encuentran todos los tratamientos considerados junto con una descripción de los parámetros de cada uno.



Temperatura

- **30 °C:** En las gráficas de caja y brazo a) y b) de la Figura 1 se encuentran los resultados para un valor de Temperatura de 30 °C. Uniendo los resultados de ambos tipos de Cultivo (Ácido y Básico) el Número final de bacterias se encuentra entre 93 y 103 unidades y tiene un valor medio de 97.65. En este caso la mediana del Número final de bacterias tuvo un incremento de 95 a 100 al pasar de un Cultivo Básico a un Cultivo Ácido.
- **60 °C:** En las gráficas de caja y brazo c) y d) de la Figura 1 se encuentran los resultados para un valor de Temperatura de 60 °C. Uniendo los resultados de ambos tipos de Cultivo (Ácido y Básico) el Número final de bacterias se encuentra entre 101 y 106 unidades y tiene un valor medio de 102.8. En este caso la mediana del Número final de bacterias no tuvo cambio al pasar de un Cultivo Básico a un Cultivo Ácido.
- **90 °C:** En las gráficas de caja y brazo e) y f) de la Figura 1 se encuentran los resultados para un valor de Temperatura de 90 °C. Uniendo los resultados de ambos tipos de Cultivo (Ácido y Básico) el Número final de bacterias se encuentra entre 88 y 100 unidades y tiene un valor medio de 94.75. En este caso la mediana del Número final de bacterias tuvo un incremento de 92.5 a 97 al pasar de un Cultivo Básico a un Cultivo Ácido.

Comentarios generales El análisis anterior sugiere que el Número final de bacterias varía dependiendo de la Temperatura en el sentido de que el Número final de bacterias tiende a ser mayor si la Temperatura es de 60 °C, menor si es de 30 °C y aún menor si es de 90 °C.

Para los niveles de Temperatura de 30 °C y 90 °C, el Número final de bacterias tiende a aumentar si el tipo de Cultivo cambia de Básico a Ácido. Esto sugiere que el tipo de Cultivo influye sobre el Número final de bacterias al menos para estos dos casos. La relación entre Cultivo y Número final de bacterias no es del todo clara para el caso de una Temperatura de 60 °C.

La situación es ambigua con respecto a los niveles de Concentración ya que, para casi todos los tratamientos, es posible encontrar observaciones con niveles de Concentración tanto Baja como Alta en ambos brazos de las gráficas de caja y brazo. Esto sugiere que el nivel de Concentración no tiene una gran influencia sobre el Número final de bacterias.

3 ¿Puede considerar que hay diferencia entre los niveles medios de temperatura? Esto es, identifica efectos sobre el número de bacterias por cambio en los niveles de temperatura?

Para responder a esta pregunta y a las siguientes cuatro, se propone el siguiente modelo de regresión lineal.

$$y = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_3 k + \beta_4 c + \beta_5 t_1 k + \beta_6 t_1 c + \beta_7 t_2 k + \beta_8 t_2 c + \beta_9 k c + \varepsilon$$

donde

- y es la variable respuesta correspondiente al Número final de bacterias.
- t_1 y t_2 son los regresores correspondientes a las variables indicadoras asociadas con el nivel de Temperatura, donde

$$(t_1, t_2) = \begin{cases} (0, 0) & \text{si Temperatura} = 30^\circ\text{C} \\ (1, 0) & \text{si Temperatura} = 60^\circ\text{C} \\ (0, 1) & \text{si Temperatura} = 90^\circ\text{C} \end{cases}$$

- k es el regresor correspondiente a la variable indicadora asociada con el tipo de Cultivo, donde

$$k = \begin{cases} 0 & \text{si Cultivo} = \text{Básico} \\ 1 & \text{si Cultivo} = \text{Ácido} \end{cases}$$

- c es el regresor correspondiente a la variable indicadora asociada con el nivel de Concentración, donde

$$c = \begin{cases} 0 & \text{si Concentración} = \text{Baja} \\ 1 & \text{si Concentración} = \text{Alta} \end{cases}$$

- ε es el término de error aleatorio.

Es importante resaltar que se pueden obtener 12 modelos diferentes como resultado de las distintas combinaciones de niveles de Temperatura, Cultivo y Concentración.

De modo que

- β_0 es el valor medio del Número final de bacterias para un nivel de Temperatura de 30 °C con un tipo de Cultivo Básico y con Concentración Baja.
- β_1 es una medida del cambio en el Número final de bacterias al pasar de una Temperatura de 30 °C a una Temperatura de 60 °C.
- β_2 es una medida del cambio en el Número final de bacterias al pasar de una Temperatura de 30 °C a una Temperatura de 90 °C. Entonces, la cantidad $\beta_2 - \beta_1$ es una medida del cambio en el Número final de bacterias al pasar de una Temperatura de 60 °C a una Temperatura de 90 °C.
- β_3 es una medida del cambio en el Número final de bacterias al pasar de un Cultivo Básico a un Cultivo Ácido cuando la Temperatura es de 30 °C.
- β_4 es una medida del cambio en el Número final de bacterias al pasar de una Concentración Baja a una Concentración Alta cuando la Temperatura es de 30 °C.
- β_5 mide la interacción entre los regresores t_1 y k . La cantidad $\beta_3 + \beta_5$ es una medida del cambio en el Número final de bacterias al pasar de un Cultivo Básico a un Cultivo Ácido cuando la Temperatura es de 60 °C.
- β_6 mide la interacción entre los regresores t_1 y c . La cantidad $\beta_4 + \beta_6$ es una medida del cambio en el Número final de bacterias al pasar de una Concentración Baja a una Concentración Alta cuando la Temperatura es de 60 °C.
- β_7 mide la interacción entre los regresores t_2 y k . La cantidad $\beta_3 + \beta_7$ es una medida del cambio en el Número final de bacterias al pasar de un Cultivo Básico a un Cultivo Ácido cuando la Temperatura es de 90 °C.
- β_8 mide la interacción entre los regresores t_2 y c . La cantidad $\beta_4 + \beta_8$ es una medida del cambio en el Número final de bacterias al pasar de una Concentración Baja a una Concentración Alta cuando la Temperatura es de 90 °C.
- β_9 mide la interacción entre los regresores k y c . La cantidad β_9 es un ajuste en el Número final de bacterias cuando el tipo de Cultivo es Ácido y la Concentración es Alta.

El ajuste obtenido para el modelo propuesto está dado por

$$\hat{y} = 94.32 + 8.65t_1 - 3.0t_2 + 5.37k + 1.77c - 5.7t_1k - 1.3t_1c - 1.6t_2k + 1.8t_2c - 0.93kc$$

Para el ajuste obtenido se calculó un valor de R^2 de 83.99 % y un valor de R^2_{ajustada} de 81.11 %. Esto quiere decir que los regresores incluidos en el modelo logran explicar una cantidad suficiente de la variabilidad de la respuesta.

A continuación se realizan algunas pruebas de hipótesis para analizar los efectos de los cambios de Temperatura sobre el Número final de bacterias.

- Para $H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$ se obtuvo un estadístico $t = 124.55$ y un valor $p = 1.11 \times 10^{-16}$, que es lo suficientemente pequeño para rechazar H_0 .
- Para $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ se obtuvo un estadístico $t = 8.51$ y un valor $p = 2.69 \times 10^{-11}$, que es lo suficientemente pequeño para rechazar H_0 .
- Para $H_0: \beta_2 = 0$ vs. $H_1: \beta_2 \neq 0$ se obtuvo un estadístico $t = -2.95$ y un valor $p = 4.78 \times 10^{-3}$, que es lo suficientemente pequeño para rechazar H_0 .

De acuerdo con los resultados de las pruebas de hipótesis anteriores, es posible afirmar que no hay evidencia de que los coeficientes asociados con los cambios de Temperatura sean negligibles y, por lo tanto, es posible concluir que los efectos sobre el Número final de bacterias por cambio en los niveles de Temperatura sí son identificables.

4 ¿Es distinta la respuesta dependiendo del cultivo empleado?

- Para $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ se obtuvo un estadístico $t = 5.6$ y un valor $p = 9.00 \times 10^{-7}$, que es lo suficientemente pequeño para rechazar H_0 .

Como rechazamos la opción que β_3 tome el valor de cero, esto implica que el tipo de cultivo que se emplea en las muestras sí afecta o sí tiene un impacto significativo en el número de bacterias encontradas.

5 ¿Consideraría que los factores interactúan?

A continuación se harán varias pruebas de hipótesis que ayudarán a discernir si los distintos factores (Temperatura, Cultivo y Concentración) interactúan entre ellos en el resultado final de número de bacterias.

- Para $H_0: \beta_5 = 0$ vs. $H_1: \beta_5 \neq 0$ se obtuvo un estadístico $t = -4.86$ y un valor $p = 1.20 \times 10^{-5}$, que es lo suficientemente pequeño para rechazar H_0 .
- Para $H_0: \beta_6 = 0$ vs. $H_1: \beta_6 \neq 0$ se obtuvo un estadístico $t = -1.11$ y un valor $p = 2.73 \times 10^{-1}$, que es lo suficientemente grande para no rechazar H_0 .
- Para $H_0: \beta_7 = 0$ vs. $H_1: \beta_7 \neq 0$ se obtuvo un estadístico $t = -1.36$ y un valor $p = 1.78 \times 10^{-1}$, que es lo suficientemente grande para no rechazar H_0 .
- Para $H_0: \beta_8 = 0$ vs. $H_1: \beta_8 \neq 0$ se obtuvo un estadístico $t = 1.53$ y un valor $p = 1.31 \times 10^{-1}$, que es lo suficientemente grande para no rechazar H_0 .
- Para $H_0: \beta_9 = 0$ vs. $H_1: \beta_9 \neq 0$ se obtuvo un estadístico $t = -0.97$ y un valor $p = 3.34 \times 10^{-1}$, que es lo suficientemente grande para no rechazar H_0 .

Como podemos observar, todos los valores p de las pruebas de hipótesis son valores grandes (mayores a 0.05), con excepción del valor de la prueba para β_5 . Sin embargo, como β_5 indica la interacción entre el cambio de 30°C a 60°C de Temperatura y el tipo de Cultivo, también tenemos que tomar en cuenta β_6 para modelar la interacción total entre estas dos variables.

6 Valide su modelo mediante el análisis de los residuales.

Realizaremos un análisis de residuales para ver si el modelo cumple con todos los supuestos: 1. Varianza constante 2. Errores independientes 3. Errores se distribuyen normal con media 0

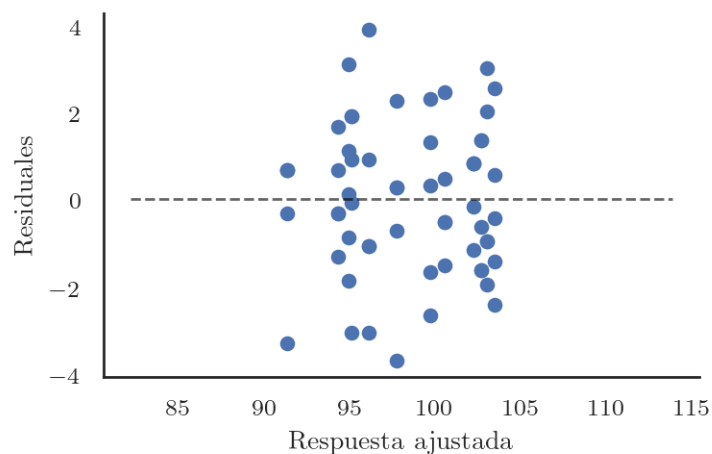


Figura 2. Residuales contra respuesta ajustada

En la figura 2 vemos que el modelo sí cumple con el supuesto de homocedasticidad pues la varianza de los residuales es prácticamente constante. Al inicio tiene una menor varianza sin embargo después se estabiliza y se mantiene constante.

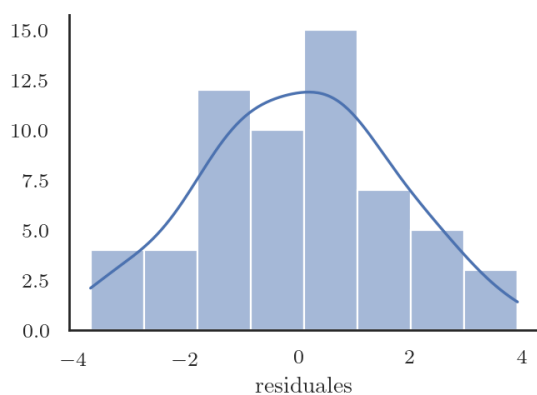


Figura 3. Histograma de los residuales

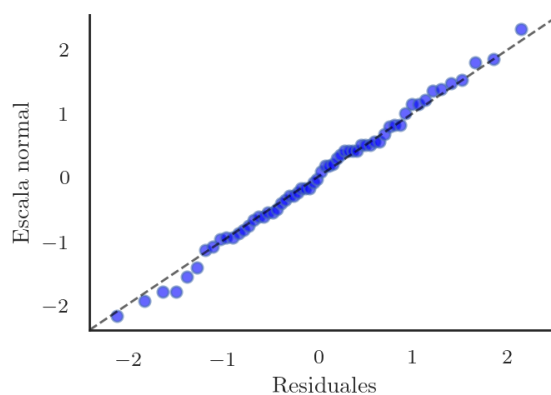


Figura 4. Gráfica cuantil-cuantil

En la Figura 3 vemos que, aunque no es idéntica, la distribución de los residuales se asemeja bastante a la de una normal. Esto se confirma con la gráfica cuantil-cuantil de la Figura 4 ya que, salvo al inicio y al final, los residuales sí se ajustan a la recta y por lo tanto cumplen con el supuesto de normalidad.

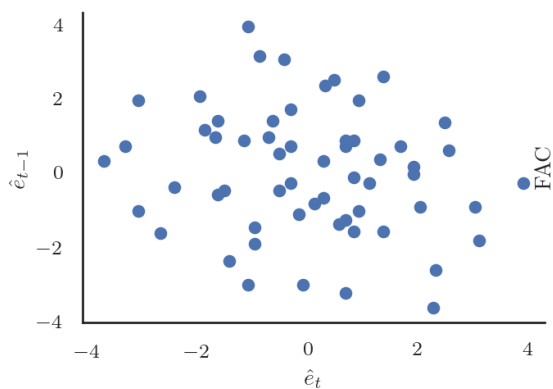


Figura 5. Gráfica \hat{e}_t vs \hat{e}_{t-1}

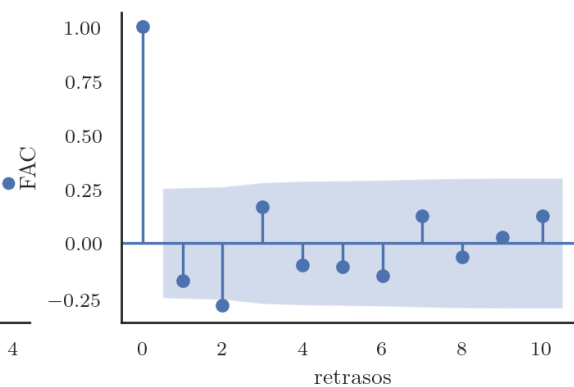


Figura 6. Función de autocorrelación

Finalmente con la Figura 5 y la Figura 6 vemos que los residuales no presentan ninguna correlación aparente entre ellos y que la correlación de orden uno es muy pequeña por lo que también se cumple el supuesto de independencia.

Una vez que nuestros residuales sí cumplen los 3 supuestos podemos afirmar que nuestro modelo es válido y útil para analizar nuestros datos.

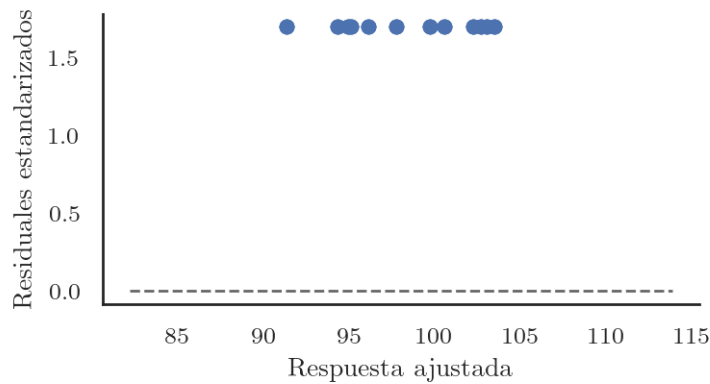


Figura 7. Residuales estandarizados contra respuesta ajustada

Finalmente de la Figura 7 vemos que los residuales estandarizados son menores que 1.75 y mayores que 0 por lo que podemos concluir que nuestra muestra no cuenta con valores atípicos.

7 ¿Considera que es necesario distinguir entre los niveles bajo y alto de la concentración del cultivo?

Para la prueba de hipótesis $H_0 : \beta_4 = 0$ vs. $H_1 : \beta_4 \neq 0$ se obtuvo un estadístico $t_4 = 1.84$ y un resultante valor $p = 7.1110 \times 10^{-2}$, que no es lo suficientemente pequeño como para rechazar la hipótesis H_0 . Esto significa que el coeficiente β_4 no es significativo y por lo tanto no es necesario distinguir entre los niveles de concentración ya que no nos brinda información extra.

8 Concluya el reporte resaltando lo encontrado en su análisis.

	df	F	PR(>F)
t1	1.0	1.6880e+02	1.1941e-17
t2	1.0	2.4443e+01	9.0172e-06
k	1.0	2.6526e+01	4.4140e-06
t1:k	1.0	2.3261e+01	1.3658e-05
t2:k	1.0	1.8601e+00	1.7872e-01
t2:c	1.0	1.3965e+01	4.7895e-04
t1:c	1.0	5.7349e-28	1.0000e+00
c	1.0	2.4559e+00	1.2339e-01
k:c	1.0	9.4943e-01	3.3456e-01
Residual	50.0	NaN	NaN

Realizamos un análisis de varianza para ver cuáles regresores podemos eliminar de nuestro modelo. Como se observó en las pruebas de hipótesis realizadas en secciones anteriores tanto la Temperatura como el tipo de Cultivo, así como su interacción, son variables importantes para determinar el número de bacterias final.

Además con este análisis observamos que la interacción entre la Temperatura y la Concentración también es un regresor importante para explicar nuestros datos ya que tiene un valor p de $4.79 \times 10^{-4} < 0.05$. Es importante notar que la prueba de hipótesis de la pregunta 5 nos indica que si la interacción entre Temperatura y Concentración ($t_2:c$) es el último regresor en agregarse al modelo, entonces no es muy significativo. Sin embargo, la tabla anterior nos indica que si solo tomamos en cuenta la Temperatura y el Cultivo en nuestro modelo, agregar el regresor $t_2:c$ sí nos ayuda a explicar de mejor manera nuestros datos. Por el *Principio de Herencia* al agregar la interacción entre Temperatura y Concentración también debemos de agregar el regresor de Concentración a la muestra aunque este no sea tan significativo.

Por lo tanto, el modelo seleccionado para explicar el número de bacterias es de la forma:

$$y = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_3 k + \beta_4 c + \beta_5 t_1 k + \beta_6 t_2 k + \beta_7 t_1 c + \beta_8 t_2 c + \varepsilon$$

y el modelo ajustado es:

$$\hat{y} = 94.55 + 8.65t_1 - 3.00t_2 + 4.90k + 1.30c - 5.70t_1k - 1.60t_2k - 1.30t_1c + 1.80t_2c$$

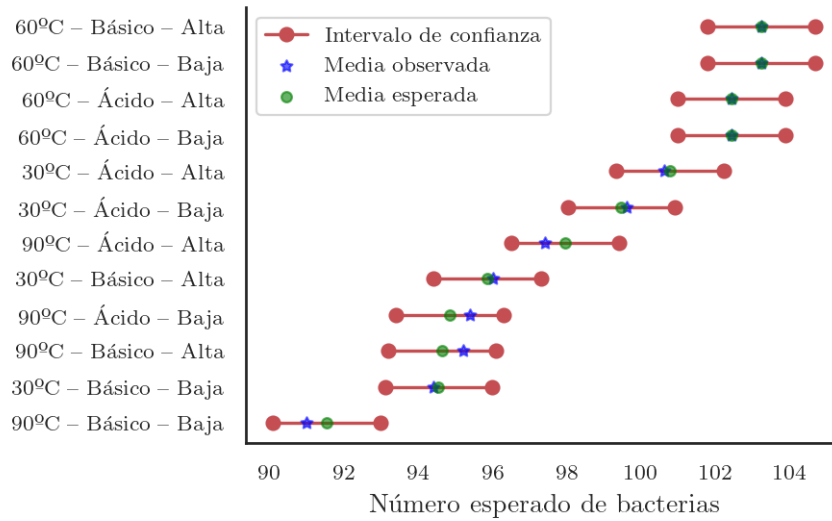


Figura 8. Intervalos de confianza del 95% para los diferentes experimentos

En la Figura 8 vemos el ajuste final de nuestro modelo el cual tiene un buen ajuste ya que las medias observadas sí están dentro de los intervalos de confianza de las predicciones del modelo.

Notamos que las bacterias reaccionan mejor a una temperatura de 60°C sin importar el tipo de cultivo ya que entre el cultivo ácido y básico hay menos de un punto de diferencia. Además en el caso de los 60°C no importa la concentración inicial del experimento pues con ambas se obtienen los mismos resultados. Por el otro lado, para temperaturas de 30°C y 90°C la bacterias crecen mejor con un cultivo ácido y concentraciones iniciales altas, pero prefieren temperaturas más bajas (30°C). Todo esto concuerda con nuestras observaciones iniciales.