

# Estadística Aplicada II - Control 3

November 7, 2019

**The PowerPuff Girls:** 154221, 156241

## 1 Introducción

Las bacterias son los organismos más abundantes del planeta. Son ubicuas, se encuentran en todos los hábitats terrestres y acuáticos; crecen hasta en los más extremos como en los manantiales de aguas calientes y ácidas, en desechos radioactivos, en las profundidades tanto del mar como de la corteza terrestre. Algunas bacterias pueden, incluso, sobrevivir en las condiciones extremas del espacio exterior. Se estima que se pueden encontrar en torno a 40 millones de células bacterianas en un gramo de tierra y un millón de células bacterianas en un mililitro de agua dulce.

En las bacterias, el aumento en el tamaño de las células (crecimiento) y la reproducción por división celular están íntimamente ligados, como en la mayor parte de los organismos unicelulares, ocurre por duplicación y se obtienen células con información hereditaria idéntica. Las bacterias crecen hasta un tamaño fijo y después se reproducen por fisión binaria, una forma de reproducción asexual.

En un laboratorio clínico se llevó a cabo un experimento para valorar cómo el nivel de temperatura y la naturaleza ácida de cierto caldo de cultivo influía en el crecimiento de una población de bacterias. Más específicamente, se consideraron dos factores: temperatura, a tres distintos niveles: 30C, 60C y 90C; dos tipos de caldo de cultivo: básico (B) y ácido (A), cada uno de ellos a dos niveles de concentraciones: baja (L) y alta (H). Como respuesta se consideró como se modificaba una población inicial de 1000 bacterias después de 48 horas de en el ambiente controlado.

A continuación se analiza cómo los distintos efectos del ambiente influyeron en el desarrollo de la población de bacterias.

## 2 Descripción de los datos

Comenzaremos el análisis de los resultados del experimento analizando gráficamente los datos, dado que el objetivo del experimento era valorar cómo el nivel de temperatura y la naturaleza ácida de cierto caldo de cultivo influía en el crecimiento de una población de bacterias, graficaremos el número de bacterias contra la acidez y contra la temperatura.

En la gráfica 1a), podemos notar que en el tipo de cultivo básico el rango del número de bacterias es más extenso por lo cuál para ambos tipos de concentración con el tipo de cultivo básico se alcanza el máximo y el mínimo. Si sólo tomáramos en cuenta la concentración, los extremos (máximos y mínimos) se alcanzan cuando la concentración es baja, sin embargo, si tomamos la mediana de los datos se tiene que la mediana de la concentración alta es más grande para ambos tipos de cultivo.

En la gráfica 1b), también observamos que el cultivo básico aumenta lo extenso del rango del número de bacterias. Notamos que si tomamos en cuenta la temperatura, para 30C y 90C el

cultivo ácido aumenta el número de bacterias. Sin embargo, para 60C la mediana en ambos tipos de cultivo es muy similar y parece no tener tanta influencia en el número de bacterias. En la gráfica 1c), se puede observar un comportamiento similar a la gráfica 1b), a concentraciones altas parece aumentar el número de bacterias si la temperatura es de 30C y 90C. pero para 60C parece no influir la concentración una vez que la media es muy similar.

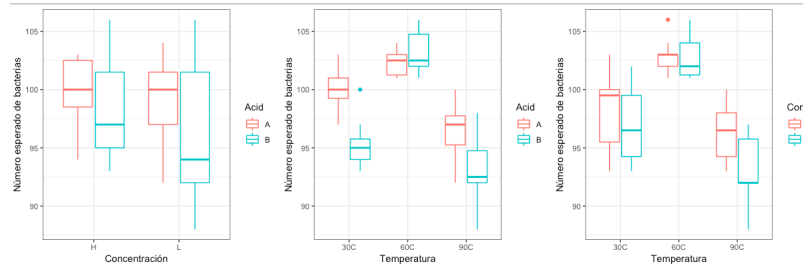


Figure 1: Gráficas de Box

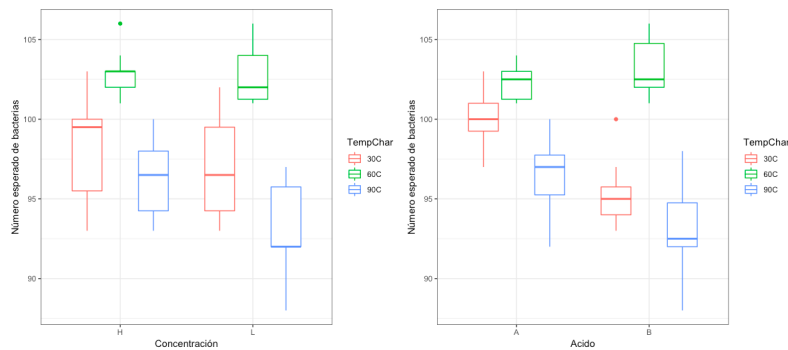


Figure 2: Gráficas de Box: Temperatura

Por otro lado, si comparamos las condiciones entre las temperaturas, podemos observar en la gráfica 2a) que a la temperatura a la cual se alcanza el mayor número de bacterias es a 60C en cualquier tipo de concentración, para 30C y 90C podemos notar que a concentraciones altas el número de bacterias es ligeramente mayor que a concentraciones bajas.

En la gráfica 2b) se obtiene un resultado similar, la temperatura a la cual se alcanza el mayor número de bacterias es a 60C en cualquier tipo de cultivo, para 30C y 90C podemos notar que en cultivos ácidos el número de bacterias es mayor que en cultivos básicos.

### 3 Respuestas 2-4

A continuación, para responder a las preguntas, graficamos las diferencias que suceden al realizar un filtro de la temperatura para 30°C, 60°C y 90°C.

Notemos que en la figura 3, el grado de concentración alto y el tipo de cultivo ácido producen un crecimiento de bacterias más alto. En cuanto a la concentración no parece haber una diferencia importante. La media es de 97.50.

Para figura 4 tenemos una media de 102.5. Se puede apreciar que un tipo de cultivo básico influye en un aumento de población, sin embargo, no hay variación respecto a la concentración. Cabe señalar que a esta temperatura es donde se alcanza el número de bacterias más alto.

En la figura 5 se nota una diferencia de resultados respecto al tipo de concentración, ya que cuando

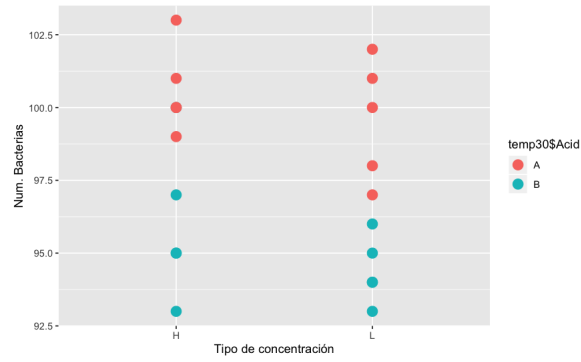


Figure 3: Temperatura 30C

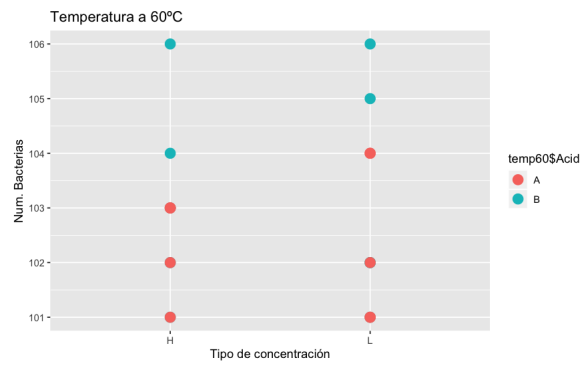


Figure 4: Temperatura 60C

es alto tenemos un número de bacterias superior a 92.5 (por encima del primer cuartil) con un valor máximo de 100. Por el otro lado, cuando es baja la concentración vamos desde 88 a 100. Notemos que cuando tenemos una acidez alta, la población tiende a ser más alta. La media es de 95.00.

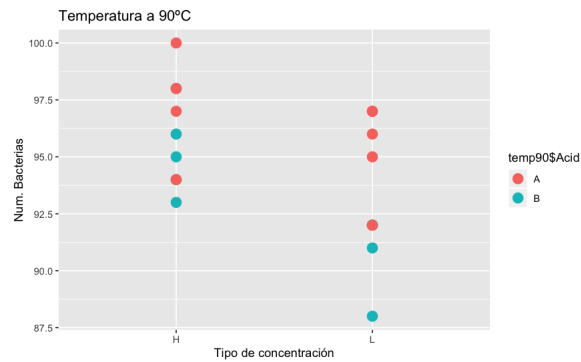


Figure 5: Temperatura 90C

Al analizar cada uno de los diferentes casos, podemos notar que hay una mejora considerable cuando la temperatura es de 60°C, ya que la media es la más alta. Además, cuando el tipo es básico y la concentración es baja, se tienen mejores resultados. Por lo anterior, podemos concluir que el tipo de cultivo sí modifica el resultado y además que los factores interactúan entre sí.

De forma más general, si analizamos y observando las gráficas de interacción de la figura 6 el número esperado de bacterias esperadas tomando en cuenta los diversos factores podemos notar que:

Para temperaturas de 30C y 90C se espera mayor número de bacterias si se tiene un caldo de cultivo de tipo ácido, en comparación con uno básico, cuando la temperatura es de 60C no hay mucha diferencia del número de bacterias esperado si el cultivo es de tipo básico o ácido.

Para concentraciones altas se espera mayor número de bacterias. Sin embargo, a temperatura de 30°C la respuesta (el número esperado de bacteria) es igual en ambos tipos de concentración.

Para ambos tipos de cultivo, a mayor concentración mayor número esperado de bacterias.

Para ambos tipos de concentración, si el cultivo es de tipo ácido se espera un mayor número de bacterias.

En particular, tomando en cuenta estas dos variables (Tipo de cultivo y Concentración) el mayor número de bacterias se alcanza cuando la concentración es alta y el tipo de cultivo es ácido. Si además, se toma en cuenta la temperatura.

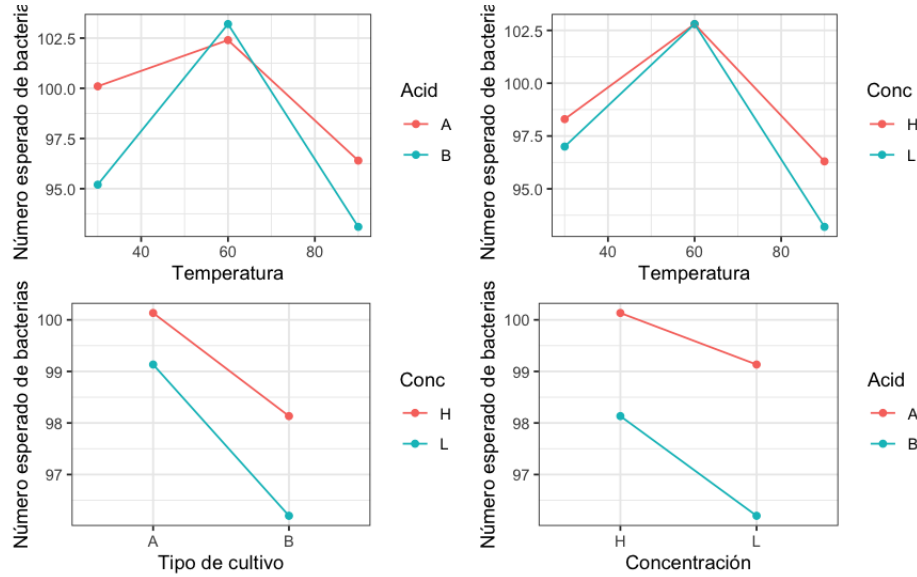


Figure 6: Gráficas de interacción

## 4 Modelo de regresión

Ahora, procederemos a realizar un primer modelo de regresión lineal que esté dado por:

$$NumBacterias = temperatura + acidez + concentración$$

Notemos que el valor de  $R^2$  es de 0.73, los factores son significativos y modelo presenta valores p pequeños y menores a 0.05. En la gráfica de residuales se nota un pequeño incremento al llegar al valor 100, pero en segundo gráfico se puede notar que los residuales estandarizados sí siguen una distribución normal.

```
Call:
lm(formula = Nbact ~ TempChar + Acid + Conc, data = datos2)

Residuals:
    Min       1Q   Median       3Q      Max
-4.783 -1.704 -0.150  1.308  5.167

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  99.6167    0.6582  151.337 < 2e-16 ***
TempChar60C   5.1500    0.7211   7.142 2.21e-09 ***
TempChar90C  -2.9000    0.7211  -4.022 0.000178 ***
AcidB        -2.4667    0.5887  -4.190 0.000102 ***
Conc         -1.4667    0.5887  -2.491 0.015779 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.28 on 55 degrees of freedom
Multiple R-squared:  0.7338,    Adjusted R-squared:  0.7145
F-statistic: 37.91 on 4 and 55 DF,  p-value: 3.293e-15
```

Figure 7: Análisis de regresión

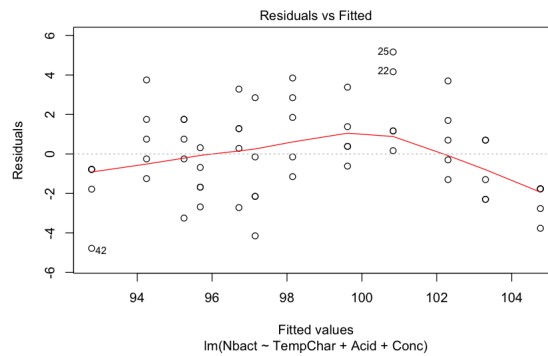


Figure 8: Residuals vs Fitted

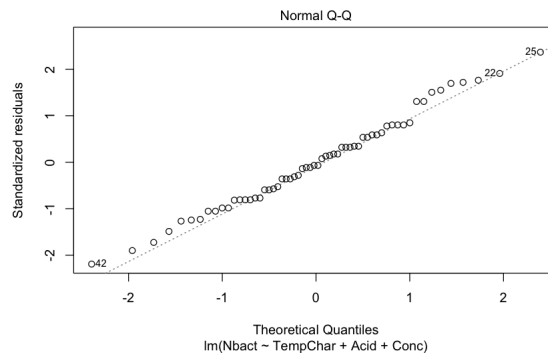


Figure 9: Normal Q-Q

En el análisis de la varianza al observar los valores-F podemos ver que la temperatura y el tipo ácido son muy importantes en el modelo. Así mismo, nos muestra que el regreso de la concentración, a pesar de que no tiene un valor comparado con el tiempo y el ácido, también aporta una parte

importante.

```

Analysis of Variance Table

Response: Nbact
      Df Sum Sq Mean Sq F value    Pr(>F)
TempChar  2 664.90   332.45  63.9401 4.471e-15 ***
Acid      1  91.27    91.27  17.5533 0.000102 ***
Conc      1  32.27    32.27   6.2059 0.015779 *
Residuals 55 285.97     5.20
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 10: ANOVA 1

## 5 ANOVA de interacción de dos vías

El análisis de varianza de dos vías, también conocido como plan factorial con dos factores, sirve para estudiar la relación entre una variable dependiente cuantitativa y dos variables independientes cualitativas (factores) cada uno con varios niveles. El ANOVA de dos vías permite estudiar cómo influyen por si solos cada uno de los factores sobre la variable dependiente (modelo aditivo) así como la influencia de las combinaciones que se pueden dar entre ellas (modelo con interacción).

A continuación estudiamos el efecto del número de bacterias (variable cuantitativa dependiente) dependiendo de la concentración (alta o baja) y de la temperatura (a los tres niveles 30C, 60C y 90C).

El efecto de la interacción doble consiste en estudiar si la influencia de uno de los factores varía dependiendo de los niveles del otro factor. Es decir, si la influencia del factor concentración sobre el número de bacterias es distinta según la temperatura o lo que es lo mismo, si el número de bacterias para determinada temperatura es distinta si se tiene una alta o baja concentración.

El modelo quedaría de la siguiente manera:

$$\text{NúmeroBacterias} = +\text{concentración} + \text{temperatura} + \text{concentración} : \text{temperatura} + \epsilon$$

Para poder dar por validos los resultados del ANOVA es necesario verificar que se satisfacen las condiciones de un ANOVA.

Primero checamos el supuesto de normalidad de los residuales del modelo, la figura 11 parece indicar que se sigue una distribución normal, a pesar de que en los extremos se tienen valores atípicos.

Para confirmar lo anterior realizamos la prueba de Shapiro que arroja un **\*\*p-value\*\*** de 0.4964 lo cual indica es probable que los residuos se distribuyan normales.

Basado en lo anterior, podemos asumir que el supuesto de normalidad de los residuos del modelo se cumple.

Checamos el segundo supuesto de homocedasticidad, es decir, homogeneidad de las varianzas. La gráfica Residuals vs Fitted de la figura 13 parece indicar que las varianzas entre los grupos son homogéneas. Por lo que se cumple el segundo supuesto.

Una vez que los supuestos se cumplen implementamos el modelo. La tabla de la figura 14 muestra que: La variable temperatura tiene un efecto significativo ( $3.82 \times 10^{-13}$ ), la concentración también (0.0305). Respecto a la interacción, se acepta la hipótesis nula (0.1665 mayor de 0.05), esto es no

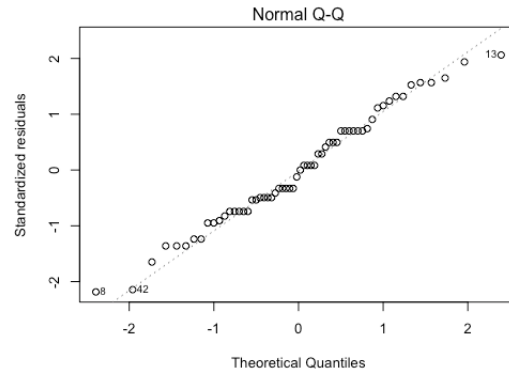


Figure 11: ANOVA 2: Normal QQ

```
##
## Shapiro-Wilk normality test
##
## data: aov_residuals
## W = 0.98153, p-value = 0.4964
```

Figure 12: ANOVA 2: Shapiro Test

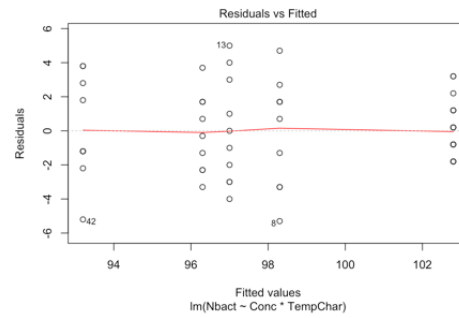


Figure 13: ANOVA 2: Residuals vs Fitted

hay interacción entre ambas variables.

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Conc      1   32.3     32.3    4.936  0.0305 *
## TempChar   2  664.9    332.5   50.856 3.82e-13 ***
## Conc:TempChar 2   24.2     12.1    1.854  0.1665
## Residuals 54  353.0      6.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 14: ANOVA 2: Resumen

## 6 Conclusión

Tras la descripción gráfica pudimos notar una fuerte interacción entre el tipo de acidez y la temperatura, pero no se veía tan marcada la relación entre la concentración y la temperatura. El análisis de los datos mostró que la temperatura óptima que maximiza el número de bacterias es de 60°C, la concentración que más favorece el número de bacterias es alta y el tipo de cultivo ácido. Al realizar la primera regresión corroboramos el hecho de que los factores con más interacción son la temperatura y la naturaleza ácida del cultivo, sin embargo, también aporta un buen nivel de significancia la concentración. Cabe recordar que obtuvimos una  $R^2$  de .73, residuales aleatorios y un en general un primer buen modelo de regresión. Para el modelo ANOVA de interacción de la sección 5, concluimos que no hay diferencias significativas en el número de bacterias según la interacción del tipo de concentración y la temperatura a la que se somete.