

## INFORMACIÓN TÉCNICA

# ¿DESPUÉS DE UN ANÁLISIS DE VARIANCIA...QUÉ? EJEMPLOS EN CIENCIA DE ALIMENTOS<sup>1</sup>

*Eric Wong-González<sup>2</sup>*

### RESUMEN

**Después de un análisis de variancia... qué? Ejemplos en ciencia de alimentos.** Es común encontrar publicaciones científicas en las que el análisis de variancia (ANDEVA) se utiliza incorrectamente y además, en las que los análisis posteriores tras encontrar significancia en el ANDEVA también son incorrectos. En este trabajo se describe el fundamento teórico del ANDEVA y de las pruebas que pueden aplicarse posteriormente, y se presentan ejemplos prácticos de la ciencia de los alimentos que tienen el objetivo de mostrar al lector la escogencia y aplicación apropiadas de estas pruebas.

**Palabras claves:** ANDEVA, comparación de medias, Dunnet, Duncan, Tukey, contrastes ortogonales.

### ABSTRACT

**What to do after an ANOVA: Examples in food science.** It is common to find scientific publications with incorrect uses of the analysis of variance (ANOVA), as well as incorrect applications of tests after finding significance with the ANOVA. This paper describes the fundamentals of the ANOVA and of the tests that can be applied afterwards, and presents practical examples in food science, with the objective of showing the reader a right selection and application of such tests.

**Key words:** ANOVA, mean comparison, Dunnet, Duncan, Tukey, orthogonal contrasts.



### INTRODUCCIÓN

En investigación existen muchos tipos de experimentos, dentro de los cuales es frecuente encontrar el experimento con un solo factor. El análisis de variancia (ANDEVA) es la técnica utilizada para interpretar los resultados de este tipo de experimento. Sin embargo, es común encontrar publicaciones en las que el ANDEVA se utiliza incorrectamente y además,

en las que los análisis posteriores tras encontrar significancia en el ANDEVA también son incorrectos. Es criterio del autor que las dos principales razones por las que esto sucede son: 1) desconocimiento de los fundamentos teóricos en los que se basa cada prueba y 2) desapego de la prueba escogida con el verdadero objetivo de la investigación. A continuación, en una primera sección, se describe el fundamento teórico del ANDEVA y de las pruebas que pueden aplicarse

<sup>1</sup> Recibido: 23 de febrero, 2010. Aceptado: 22 de noviembre, 2010.

<sup>2</sup> Escuela de Tecnología de Alimentos, Centro Nacional de Ciencia y Tecnología de Alimentos, Universidad de Costa Rica, San José, Costa Rica. eric.wong@ucr.ac.cr

posteriormente. En una segunda sección se presentan ejemplos prácticos que tienen el objetivo de mostrar al lector la escogencia y aplicación apropiadas de estas pruebas. Los ejemplos se basan en experimentos de un solo factor, correspondientes a diseños irrestrictos aleatorios con mediciones independientes. Los datos no necesariamente corresponden a experimentos reales y pueden haber sido modificados para lograr mostrar lo requerido.

## EL ANDEVA Y PRUEBAS POSTERIORES

El análisis de variancia es una técnica que se usa para probar hipótesis acerca de las medias de diferentes tratamientos que se ensayan (Daniel 1984). Para que esto sea posible, las medias que se determinan tienen que provenir de mediciones sobre muestras independientes, pues si las mediciones se hacen sobre la misma muestra deben utilizarse técnicas que contemplen mediciones repetidas. Además, el ANDEVA no puede utilizarse para porcentajes o proporciones de naturaleza binomial (como por ejemplo porcentaje de células muertas), pues éstos no cumplen el supuesto de igualdad de variancias. Por el contrario, el ANDEVA sí aplica para porcentajes que respondan a una distribución normal (como por ejemplo % de humedad de un alimento) (Kuelh 2000, Montgomery 2003).

Cuando el ANDEVA se utiliza para analizar los datos de un experimento de un solo factor, como los que se discuten en este documento, la primera etapa es comprobar si hay significancia del factor (tratamientos) que se estudió. Si no hay significancia, un análisis posterior es innecesario. De acuerdo con González (2008), procede reportar la no significancia e indicar que no hay evidencia para concluir que los tratamientos ensayados provocan diferencia en el promedio observado para la variable de interés. Junto con la no significancia, debe reportarse la potencia de la prueba ( $1-\beta$ ) (para más detalle acerca de la potencia de prueba puede consultar diversos textos de estadística o la publicación de González (2008)).

En estudios de un factor, si se comprueba la significancia de los tratamientos (niveles del factor) tras aplicar el ANDEVA, se sabría entonces que al menos uno de los promedios de la variable respuesta determinado para un tratamiento, es diferente de los

obtenidos para los otros tratamientos (Montgomery 2003). Para identificar cuál o cuáles promedios son diferentes resulta necesario realizar pruebas adicionales. Por ejemplo, puede ser que todos los promedios sean diferentes entre sí, o que haya dos grupos que se diferencian, o que solamente uno se diferencie de los demás y así sucesivamente (Navarro 2006).

Cuando se quiere comparar los promedios más a fondo, tras un ANDEVA para un experimento de un solo factor, existen tres posibilidades generales: aplicar una prueba de comparaciones múltiples, la prueba de contrastes ortogonales o un análisis de regresión. Las dos primeras tienen aplicación cuando los tratamientos que se comparan son cualitativos, mientras que en la regresión, cuando el objetivo es estudiar una tendencia, los tratamientos son de naturaleza cuantitativa.

### Pruebas de comparaciones múltiples

Existe una diversidad de pruebas de comparaciones múltiples, y resulta muy importante conocer su fundamento teórico para saber cuándo es apropiado aplicarlas. De acuerdo con Gacula y Singh (1984), entre las pruebas más conocidas están: diferencia mínima significativa (LSD por sus siglas en inglés), Dunnet, Duncan, Tukey, Newman-Keuls y Scheffé. Este mismo autor indica que no necesariamente se obtiene el mismo resultado al aplicar todas estas pruebas a un mismo conjunto de datos. Esto demuestra la importancia de entender en qué casos es aplicable cada una. Adicionalmente, es importante comprender que el carácter múltiple de las comparaciones implica una desventaja para controlar el Error  $\alpha$  (Tipo I). Se sugiere al lector interesado ahondar en este tema con base en las referencias suministradas.

A continuación se describen las tres pruebas de comparaciones múltiples con mayores diferencias entre sí, y que además, son las de frecuente aplicación en el campo de investigación, pues son consideradas por diversos autores como más confiables que otras similares en una determinada aplicación (por ejemplo LSD versus Tukey en un mismo conjunto de datos).

#### *Prueba de Dunnet*

Esta prueba se utiliza cuando el objetivo de la investigación es comparar las medias de todos los tratamientos contra un control, el cual a su vez, se

considera como un tratamiento. Esta prueba no permite establecer diferencias entre las medias de los otros tratamientos entre sí. Montgomery (2003) y Gacula y Singh (1984) discuten ampliamente el procedimiento de cálculo para esta prueba.

#### *Prueba de Duncan*

La prueba de Duncan, también conocida como la prueba de rango múltiple, es conveniente aplicarla cuando los tamaños de las muestras son iguales y los tratamientos presentan una relación ordinal, es decir, pueden ordenarse de manera ascendente o descendente en una escala no numérica, a diferencia de los tratamientos que responden a variables continuas en las que procede un análisis de regresión. Navarro (2006) y Montgomery (2003) presentan las fórmulas de cálculo y particularidades de esta prueba.

#### *Prueba de Tukey*

La prueba de Tukey es la prueba más aplicada y preferida por los estadísticos, pues controla de mejor manera los dos errores ampliamente conocidos en la estadística ( $\alpha$  y  $\beta$ ) (Montgomery 2003). Esta prueba permite hacer todas las posibles comparaciones de tratamientos de dos en dos, y por eso se considera la más completa de las tres pruebas aquí descritas. Wu y Hamada (2000), discuten en detalle el procedimiento para aplicar esta prueba.

### **Contrastes ortogonales**

Todos los autores hasta ahora citados en este trabajo discuten la pertinencia del uso de contrastes ortogonales como una mejor alternativa para sustituir las pruebas de comparaciones múltiples antes descritas. Todos ellos hacen una descripción muy completa de la teoría relativa a esta prueba. Navarro (2006), de una manera muy acertada y simple, indica que la prueba de contrastes ortogonales muestra mayor flexibilidad, pues permite comparaciones entre medias individuales o entre grupos de medias, sobre todo, cuando se tiene algún criterio que permite separar los tratamientos en grupos lógicos. Esto último es el requisito indispensable para poder aplicar esta prueba. La ventaja de usar contrastes ortogonales, a diferencia de la prueba

de Tukey, es que con éstos se logra un control total al estimar los errores  $\alpha$  y  $\beta$  (Montgomery 2003).

### **Análisis de regresión**

El análisis de regresión se aplica cuando se desea conocer la relación entre dos variables continuas (Samuels 1989). Así, cuando los tratamientos corresponden a niveles de una variable continua, y es objetivo de la investigación evaluar la tendencia correspondiente, se aplica un análisis de regresión. En este caso, para poder interpretar la significancia, se recomienda la valoración de dos índices estadísticos para experimentos de un solo factor: coeficiente de regresión del ajuste escogido ( $\beta_1$ ) y el coeficiente de determinación ( $R^2$ ). El ajuste que se escoja debe responder a lo que reporta la literatura para la variable respuesta en estudio (por ejemplo ajuste lineal, cuadrático, etc.). Dependiendo del ajuste así será el número de coeficientes de regresión cuya significancia se valora para determinar si hay una tendencia. Si alguno de los  $\beta$  es significativo puede decirse que hay una tendencia (Montgomery 2003). Por otra parte, el  $R^2$  indica qué porcentaje del comportamiento real de los datos es explicado por el ajuste encontrado o la tendencia obtenida (Navarro 2006). El  $R^2$  se interpreta considerando el fenómeno que se estudia, pues en algunos con poca variabilidad, éste debe ser un porcentaje alto, mientras que para otros, un 60% de explicación, por ejemplo, puede considerarse suficiente para un ajuste dado. El lector puede consultar a los autores aquí citados para más detalles acerca del análisis de regresión.

## **APLICACIONES DEL ANDEVA Y PRUEBAS POSTERIORES**

### **Ejemplo 1: Comparaciones múltiples con Dunnett**

**Descripción del problema:** Una empresa productora de café utiliza a catadores para medir la calidad de la bebida. La empresa cuenta con un maestro catador cuya experiencia de años lo convierte en una base de referencia. Existe también un grupo de cinco catadores en formación, cuyo desempeño se compara constantemente contra el del maestro catador, con el fin de decidir si pueden seguir en formación hasta alcanzar la

categoría de maestro, o si se prescinde de sus servicios. En este proceso, se hacen pruebas en las que tanto el maestro catador como los catadores en formación evalúan una muestra estándar. Los resultados obtenidos por los catadores en formación se comparan con el del maestro catador y se registran los resultados.

**Objetivo:** Comparar la calificación otorgada por cada catador en formación con la calificación otorgada por el maestro.

**Escogencia del análisis:** La calificación otorgada a la bebida por un catador es un valor puntual y no tiene relación con una tendencia, por lo que no aplica un análisis de regresión. Es evidente que no hay interés de comparar el desempeño de los catadores en formación entre ellos, pues dicha información no tiene valor para la empresa, pues lo que interesa es saber cómo se desempeña cada uno en comparación con el maestro catador. Esto descarta las pruebas de contrastes ortogonales y la de Tukey. Por otro lado, no hay una relación de orden entre los catadores en formación, ni interesa establecerla, por lo que la prueba de Duncan no ofrece alguna utilidad en este caso. Dado el objetivo de la empresa y la descripción del problema, la prueba más adecuada para realizar el análisis de los resultados es la de Dunnet, al visualizar fácilmente un tratamiento control (maestro catador) y la necesidad de comparar los tratamientos (catadores en formación) con dicho control.

**Resultados:** Tras aplicar el ANDEVA (para un  $\alpha=0,05$ ) se obtuvo significancia para el factor que se denominó "tipo de analista" ( $p<0,0001$ ) e indica que al menos una de las calificaciones promedio es diferente de las otras. En el Cuadro 1 se muestran los promedios para las calificaciones otorgadas a la muestra estándar por el maestro y los catadores en formación, con letras diferentes se señala cuáles promedios son diferentes del promedio asignado a la bebida por parte del maestro catador.

La prueba de Dunnet permite entonces concluir, a un nivel de significancia de 0,05, que solamente los catadores en formación 1 y 4 aprobaron la prueba, pues su calificación promedio es igual a la del maestro catador, mientras que los catadores en formación 2, 3 y 5 asignaron calificaciones promedio diferentes a la asignada por el maestro catador.

## Ejemplo 2: Comparaciones múltiples con Duncan

**Descripción del problema:** Se quiere determinar el efecto del nivel de energía de un microondas sobre

**Cuadro 1.** Calificación promedio asignada a la bebida estándar de café según analista.

Analista	Calificación promedio <sup>1</sup>
Maestro catador	91 ± 1 a
Catador en formación # 1	91 ± 1 a
Catador en formación # 2	88 ± 1 b
Catador en formación # 3	66 ± 3 b
Catador en formación # 4	91 ± 1 a
Catador en formación # 5	78 ± 2 b

<sup>1</sup>Las calificaciones promedio con letra diferente al control (maestro catador) son diferentes al control, según la prueba de Dunnet ( $\alpha=0,05$ ). Los promedios se muestran con el límite de confianza al 95%.

el tiempo requerido para llevar un litro de agua a ebullición. El microondas disponible en el laboratorio tiene un selector para el nivel de energía con tres niveles: baja, media y alta. Estos niveles corresponden a un ámbito de energías y no a una energía puntual, producto de las variaciones normales que presenta el horno en su funcionamiento.

**Objetivo:** Determinar si el nivel de energía de un microondas tiene efecto sobre el tiempo requerido para llevar un litro de agua a ebullición.

**Escogencia del análisis:** De acuerdo con el objetivo se requiere comparar los promedios de tiempo encontrados para cada nivel de energía. No hay un requerimiento de evaluación de tendencias por lo que no aplica el análisis de regresión. Los tratamientos (tres niveles de energía) no pueden agruparse entre sí con algún sentido teórico, por lo que no procede analizar contrastes ortogonales. No existe, ni tiene sentido para el experimento, la evaluación de un control, lo que descarta la prueba de Dunnet. La prueba de Tukey haría la comparación respectiva pero sin considerar la relación ordinal que existe entre los tratamientos. Por lo anterior, la prueba más adecuada para este ejemplo es la de Duncan.

**Resultados:** El "nivel de energía" resultó un factor significativo ( $p<0,0001$ ) tal y como lo evidenció el ANDEVA (para un  $\alpha=0,05$ ), indica que al menos uno de los tiempos requeridos para ebullición de un litro de agua fue diferente a los otros. El tiempo que se requiere para llevar el agua a ebullición en el nivel de energía alto es significativamente menor ( $p<0,05$ ) que los

tiempos requeridos a los niveles de energía medio y bajo (Cuadro 2). Estos dos últimos niveles de energía no causan tiempos de ebullición diferentes ( $p > 0,05$ ). En términos prácticos la prueba de Duncan nos permite concluir que si un tiempo de 7,3 minutos es adecuado para nuestros intereses, basta con aplicar un nivel de energía bajo, pues el medio no nos ofrecería un proceso más rápido. Por el contrario, si ocupamos el proceso más rápido posible, la prueba de Duncan nos ayuda a saber que debemos aplicar el nivel de energía alto.

**Cuadro 2.** Tiempo requerido para llevar un litro de agua a ebullición según nivel de energía del horno microondas.

Nivel de energía	Tiempo requerido para ebullición (min.) <sup>1,2</sup>
Bajo	7,3 ± 0,5 a <sup>▲</sup>
Medio	5,8 ± 0,3 a <sup>●</sup>
Alto	2,5 ± 0,3 b <sup>■</sup>

<sup>1</sup> Los tiempos promedio con letra diferente son significativamente diferentes, según la prueba de Duncan ( $\alpha=0,05$ ). Los promedios se muestran con el límite de confianza al 95%.

<sup>2</sup> Los tiempos promedio con símbolos diferentes son significativamente diferentes, según la prueba de Tukey ( $\alpha=0,05$ ).

En este ejemplo, resulta importante recordar que la prueba de Duncan ofrece la ventaja de considerar la relación ordinal que existe entre los tratamientos, por lo que representa una forma más pertinente de analizar los resultados en comparación con utilizar la prueba de Tukey. Sin embargo, la aplicación de la prueba de Tukey a este conjunto de datos sería también correcta. Como se puede observar en el Cuadro 2, para la prueba de Tukey los tres promedios son diferentes entre sí, por lo que la conclusión a la que se llega es diferente. Con esta prueba se concluiría que el tiempo requerido para ebullición el agua es mayor para el nivel energía bajo, y que éste se reduce al aumentar la energía a nivel medio, y se reduce aún más al aumentar a alta energía. Esto demuestra que es esencial la escogencia de una prueba más pertinente, de acuerdo con los objetivos y la naturaleza de los tratamientos.

### Ejemplo 3: Comparaciones múltiples con Tukey

**Descripción del problema:** En las etapas preliminares para montar una nueva y pequeña empresa de papas tostadas resulta difícil escoger el pelador de papas que utilizarán los empleados. La variable determinante, según la opinión de expertos, es la velocidad de pelado (expresada como número de papas del mismo tamaño y forma que se pelan en un minuto). Se decide entonces realizar un experimento para determinar la velocidad de tres diferentes peladores de papas que el nuevo negocio estaría en posibilidad de comprar: pelador marca A, pelador marca B y pelador marca C.

**Objetivo:** Determinar si existen diferencias en el número de papas promedio que se pelan en un minuto con cada uno de los tres peladores.

**Escogencia del análisis:** De acuerdo con el objetivo, de inmediato se reconoce la necesidad de comparar promedios. No aplica un análisis de regresión porque no se quiere estudiar una tendencia. La pregunta entonces es: ¿comparaciones múltiples o contrastes ortogonales? Para aplicar contrastes ortogonales tendría que haber al menos dos peladores de papas con una característica común que permitiera agrupar dos para contrastar con el otro, hecho que no pareciera darse en este caso (pues se trata de diferentes marcas comerciales) y que, además, no está expresado en el objetivo. Entonces es necesario definir cuál de los tres métodos de comparaciones múltiples procede aplicar. Ante la ausencia de un tratamiento control no procede aplicar Dunnett. Al no existir una relación de orden entre los peladores no procede aplicar Duncan. Por lo tanto, la comparación adecuada en este caso es Tukey.

**Resultados:** Luego de aplicar el ANDEVA y para un nivel de significancia del 5% se encontró que el efecto de “tipo de pelador” fue significativo ( $p=0,0016$ ), indicando que al menos uno de los promedios del número de papas peladas en un minuto es diferente de los otros. En el Cuadro 3 se muestran los promedios para el número de papas peladas en un minuto según el tipo de pelador. No hubo diferencia en el número de papas que se logran pelar en un minuto con el pelador A (cuatro papas) y el pelador B (cinco papas) (Cuadro 3). Sin embargo, cuando se pelan las papas con el pelador C, se logra pelar un número mayor de papas (9) en comparación con cada uno de los otros dos peladores. La aplicación de la prueba de Tukey en

**Cuadro 3.** Número de papas peladas en un minuto según tipo de pelador.

Tipo de pelador	Número promedio de papas <sup>1</sup>
Marca A	4 ± 2 a
Marca B	5 ± 1 a
Marca C	9 ± 1 b

<sup>1</sup>El número promedio de papas con letra diferente es significativamente diferente, según la prueba de Tukey ( $\alpha=0,05$ ). Los promedios se muestran con el límite de confianza al 95%.

este caso permitió decidir que la inversión más apropiada y segura son los peladores de la marca C.

#### Ejemplo 4: Contrastes ortogonales

**Descripción del problema:** En una planta procesadora de alimentos se requiere escoger un desinfectante para aplicar sobre los pisos de la zona de recepción de materia prima. Varios proveedores luchan por ser los elegidos y ofrecen su producto como el mejor. La empresa decide diseñar un experimento para comparar el efecto en el número de bacterias de *E. coli* que ofrecen los cuatro compuestos que le están ofreciendo las diferentes casas comerciales. Dos compuestos son a base de cloro y los otros dos a base de amonio cuaternario. Los desinfectantes clorados son mucho más baratos por lo que interesa compararlos contra los de amonio cuaternario. También interesa conocer si existen diferencias en las reducciones entre los dos compuestos que basan su acción en el mismo ingrediente activo. Además, la empresa quiere comprobar si aplicar los desinfectantes representa alguna ventaja con respecto a la simple aplicación de agua.

**Objetivos:** Determinar si:

- Existen diferencias entre la aplicación de los desinfectantes y la simple aplicación de agua.
- Las reducciones al aplicar compuestos a base de cloro se diferencian de las obtenidas al aplicar compuestos a base de amonio cuaternario.
- Existe diferencia en las reducciones obtenidas al aplicar compuestos a base del mismo ingrediente activo.

**Escogencia del análisis:** Se reconoce nuevamente la necesidad de comparar promedios y la ausencia de interés por evaluar una tendencia, lo que elimina un

análisis de regresión. De acuerdo con los objetivos y el planteamiento del problema se visualizan claramente grupos de tratamientos que tienen características en común. Esto indica la pertinencia de utilizar un análisis de contrastes ortogonales, ya que una prueba de Tukey no permite escoger los grupos a comparar y considera la totalidad de los tratamientos para dar el resultado. La prueba de Dunnett podría utilizarse para comparar a todos los desinfectantes contra el control, pero no permitiría lograr una conclusión para los demás objetivos planteados, por lo que sólo sería una solución parcial en este caso.

**Resultados:** El efecto de “desinfectante” (para un  $\alpha=0,05$ ) resultó significativo tras el ANDEVA ( $p=0,0036$ ) indicando que al menos una de las reducciones observadas fue distinta de las otras. Los desinfectantes, como grupo, ofrecen reducciones mayores a la reducción observada para el agua ( $p=0,0052$ ) indicando que su uso es pertinente (Contraste # 1) (Cuadro 4). Por otro lado, los dos desinfectantes a base de amonio cuaternario como grupo, ofrecen reducciones mayores a las observadas para el grupo conformado por los desinfectantes a base de cloro ( $p=0,0038$ ) (Contraste # 2). Por último, se determinó que no hay diferencia significativa en las reducciones observadas para los dos desinfectantes a base de cloro ( $p=0,4678$ ) (Contraste # 3) y lo mismo para los dos desinfectantes a base de amonio cuaternario ( $p=0,0506$ ) (Contraste # 4). Este análisis permitió concluir que, en términos de la reducción microbiana obtenida (ver promedio en el Cuadro 5), lo mejor es utilizar cualquiera de los dos compuestos comerciales a base de amonio cuaternario.

Para ilustrar el alcance limitado de la prueba de Tukey en este ejemplo, en el Cuadro 5 se muestra

**Cuadro 4.** Probabilidad de error (significancia) según contraste analizado.

Contraste	Significancia
Agua versus desinfectantes	0,0052
Desinfectantes a base de cloro versus desinfectantes a base de amonio cuaternario	0,0038
Desinfectante a base de cloro # 1 versus desinfectante a base de cloro # 2	0,4678
Desinfectante a base de amonio cuaternario # 1 versus desinfectante a base de amonio cuaternario # 2	0,0506

**Cuadro 5.** Reducción logarítmica promedio de *E.coli* ( $\log_{10}$  ufc/cm<sup>2</sup>) en piso según desinfectante.

Desinfectante	Reducción promedio ( $\log_{10}$ Ufc/cm <sup>2</sup> ) <sup>1</sup>
Agua	0,5 ± 0,3 a
A base de cloro # 1	1,0 ± 0,3 a
A base de cloro # 2	0,8 ± 0,4 a
A base de amonio cuaternario # 1	1,9 ± 0,5 b
A base de amonio cuaternario # 2	1,3 ± 0,2 ab

<sup>1</sup>Reducciones promedio con letra diferente es significativamente diferente, según la prueba de Tukey ( $\alpha=0,05$ ). Los promedios se muestran con el límite de confianza al 95%.

el agrupamiento obtenido tras aplicar dicha prueba. Obsérvese como la única diferencia clara es que la reducción de uno de los dos desinfectantes a base de amonio cuaternario ( $1,87 \log_{10}$  ufc/cm<sup>2</sup>) es mayor que las reducciones observadas para los desinfectantes clorados y el agua. Esta prueba no logra ubicar al otro desinfectante a base de amonio cuaternario en un grupo específico. Por otro lado, no se detecta diferencia entre los desinfectantes clorados y el agua, ni entre uno de los desinfectantes a base de amonio cuaternario y los dos desinfectantes clorados. Como es consenso entre diversos autores, cuando existen grupos que pueden definirse con un sustento teórico, el análisis de contrastes ortogonales resulta más potente que las pruebas de comparaciones múltiples.

### Ejemplo 5: Regresión

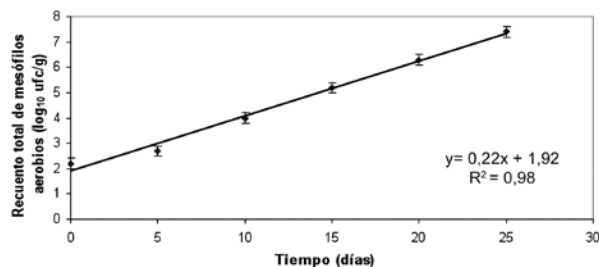
**Descripción del problema:** Una empresa trabaja con un producto en polvo y desea determinar si, durante el almacenamiento, las bacterias mesófilas aerobias se multiplican, lo cual sería negativo para la vida útil y calidad del producto.

**Objetivo:** Evaluar el comportamiento del recuento total de bacterias mesófilas aerobias durante 25 días de almacenamiento.

**Escogencia del análisis:** Si se analiza el objetivo con cuidado resulta claro que, en este caso, interesa evaluar la tendencia en el tiempo para la variable respuesta escogida (recuento total de bacterias mesófilas aerobias). Nótese que se trata de la relación entre dos

variables continuas. Pensar en una comparación múltiple de promedios (como Tukey) carece de sentido, pues no interesa determinar, por ejemplo, si el recuento a los 10 días es diferente del éste a los 13 días. La misma consideración aplica para contrastes ortogonales. Lo que interesa es determinar si la variable respuesta aumenta o no en el tiempo de estudio, para lo que se utiliza el análisis de regresión que permite determinar la significancia y magnitud del efecto del tiempo. La comunidad científica coincide en que el logaritmo del recuento de bacterias, cuando ocurre multiplicación, se comporta de manera lineal en el tiempo. Por lo tanto, corresponde evaluar la significancia del coeficiente de regresión ( $\beta_1$ ) (pendiente del ajuste lineal entre el tiempo y el logaritmo del recuento microbiano). Además, resulta importante determinar el coeficiente de determinación  $R^2$ , que nos indicaría cuán bueno es el ajuste lineal.

**Resultados:** La significancia obtenida tras el ANDEVA para el efecto tiempo fue  $p < 0,0001$  indica la existencia de una tendencia significativa. Dicha significancia es provocada por un aumento importante del recuento total aerobio en el tiempo (Figura 1). Esta variable aumentó de  $2,2 \log_{10}$  ufc/g en el día de inicio del estudio a  $7,4 \log_{10}$  ufc/g en el día final (día 25), evidenciado por una pendiente positiva ( $\beta_1=0,22$ ) y significativa ( $p < 0,0001$ ). El coeficiente de determinación  $R^2=0,98$  para el ajuste lineal (que corresponde al ajuste teórico aceptado para el fenómeno) indica que el 98% de los cambios en el recuento total de mesófilas aerobias son provocados por el paso del tiempo. Para un fenómeno microbiológico este porcentaje de variancia explicada se considera elevado, indicando un ajuste lineal muy bueno. Se concluye por lo tanto que el recuento total de mesófilos aerobios aumenta



**Figura 1.** Comportamiento del recuento total de mesófilas aerobias durante 25 días de almacenamiento de un producto en polvo.

durante el almacenamiento del producto en polvo, en detrimento de su calidad y vida útil.

### **Ejemplo 6: Sin análisis posterior**

Para cualquiera de los ejemplos analizados anteriormente, la aplicación del análisis posterior a la realización del ANDEVA asume significancia del efecto simple que estaba en estudio. Sin embargo, debe recordarse que el efecto que se estudia podría dar no significativo, por lo que no es necesario realizar un análisis posterior. En este caso lo que procede es reportar e interpretar la no significancia, junto con la potencia de prueba (González 2008). Esto debe hacerse siempre ante la no significancia del efecto simple tras el ANDEVA, independientemente de cuál método posterior de análisis se haya definido a priori.

## **COMENTARIO FINAL**

La selección de un ANDEVA para realizar un análisis de resultados merece especial cuidado para comprobar que se cumplen los supuestos respectivos y que es apropiado su uso. Una vez determinada la significancia en un experimento de un solo factor utilizando un ANDEVA, es importante conocer el sustento teórico de las posibles pruebas adicionales que pueden aplicarse para

entender la significancia encontrada. Si bien puede no haber consenso entre autores de cuál es la mejor prueba para un determinado análisis, si es consenso entre todos ellos que los objetivos del estudio determinan en gran medida cuál es la prueba más apropiada.

## **LITERATURA CITADA**

- Daniel, WW. 1984. Bioestadística. Limusa, México D.F. p. 485.
- González, MI. 2008. Potencia de prueba: la gran ausente en muchos trabajos científicos. *Agronomía Mesoamericana* 19(2):309-313.
- Gacula, MC; Singh, J. 1984. *Statistical methods in food and consumer research*. Academic Press Inc., Orlando. p. 505.
- Kuehl, RO. 2000. *Diseño de experimentos*. 2 ed. Thomson Learning, México D.F. p. 666.
- Montgomery, DC. 2003. *Diseño y análisis de experimentos*. Limusa Wiley, México D.F. p. 686.
- Navarro, JR. 2006. *Diseño experimental: aplicaciones en agricultura*. Editorial UCR, San José. p.327.
- Samuels, ML. 1989. *Statistics for the life sciences*. Prentice Hall Inc., New Jersey. p. 597.
- Wu, JCF; Hamada, M. 2000. *Experiments: Planning, analysis and parameter design optimization*. Wiley & Sons, New York. p. 630.