

# Un conocimiento básico de la estadística es necesario para los nutrólogos

Paul B. Tillman

(*Feedstuffs*, 60: 27, 14. 1988)

La estadística es la ciencia y el arte de obtener, analizar e interpretar los datos. Sus aplicaciones pueden hallarse en todas partes -Urguhart, 1971-. De ahí la absoluta necesidad de la misma para quien quiera interpretar la terminología utilizada en la descripción de las experiencias.

Los dos tipos más básicos de estadísticas que pueden aplicarse a una serie de datos son los siguientes:

1. Las medias de una tendencia central -el modo, la mediana y la media.
2. Las medidas de la dispersión -la gama, la varianza y la desviación estándar o SD.

En tanto que no nos entretengamos en la definición del modo, la mediana y la media, que pueden hallarse en cualquier libro básico de estadísticas, diremos que la gama es la diferencia entre los valores mayores y menores de una serie de datos, es decir, una medida bruta de la dispersión. Unas mediciones más útiles de la dispersión o de la variabilidad son la varianza y la SD. Aquélla es la media del cuadrado de las desviaciones, para cada observación, de la media. La SD es la raíz cuadrada de la varianza y es el indicador más frecuentemente utilizado de la dispersión: cuanto mayor es, más variabilidad existe entre los datos. Con el fin de describir perfectamente una serie de datos se requieren tanto la media, como medida de la tendencia central y la SD, para la dispersión. La media es el valor que representa al conjunto de datos y la SD nos indica el grado de dispersión entre tales datos.

Una gran proporción de datos del mundo de la biología caen dentro de una distri-

bución "normal". En este caso, representados gráficamente, adoptan una forma de campana, con un punto central, que es el modo, la mediana o la media. La zona debajo de la curva con 1 SD más cubre el 34,13% de la superficie total, ocurriendo otro tanto con la zona con 1 SD menos; por tanto, la zona cubierta de esta forma en total cubrirá el 69,26% de la superficie total. Y repitiendo lo mismo pero tomando 2 SD en más o en menos, la superficie cubierta será del 95,44%.

Ambos porcentajes pueden ser indicados en forma decimal y utilizados como valores de probabilidad -son los valores "p"-. Por tanto, utilizando la media y las SD de una serie de datos podrá determinarse el porcentaje de valores permaneciendo por arriba o por abajo de cualquier valor dentro de ellos.

La recogida de una muestra representativa y al azar de una población determinada es muy importante de cara a asegurar unos resultados precisos y exactos. Los errores en una serie de datos pueden provenir o bien de haber tomado mal las mediciones o bien del muestreo. El objetivo que debe perseguirse en una muestra al azar es el de minimizar los errores de muestreo. Las medias de una serie de muestras de una población determinada se distribuyen normalmente sobre la media de tal población. Las SD de estas medias es el "error estándar de la media" -SEM- y tiene entonces una variabilidad proveniente tanto de la medición como de los errores del muestreo. El SEM es por tanto una distribución de las medias de la muestra y puede utilizarse para sacar conclusiones sobre la misma.

# LOHMANN BROWN

Cuando se trata de huevos marrones, la Lohmann Brown ocupa una posición leader en el mercado mundial. Excelente producción de huevos, económica conversión de pienso, óptimo peso huevo, alta calidad interna y una cáscara muy uniforme – son las características más resaltantes de la Lohmann Brown.



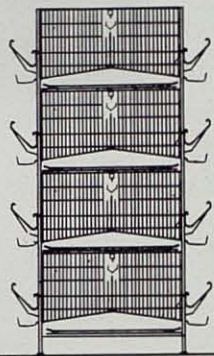
LOHMANN TIERZUCHT GMBH  
AM SEEDEICH 9-11 · D-2190 CUXHAVEN  
(ALEMANIA OCCIDENTAL)  
TEL. (47 21) 50 50  
TELEX 2 32 234 · TELEFAX (47 21) 3 24 86



# ¿Por qué hemos elegido TREVIRA ?

Porque el éxito de la jaula  
está en la calidad de la  
cinta de limpieza.

JAUJA VERTICAL



PARA CRIA-RECRÍA Y PUESTA



INDUSTRIAL GANADERA NAVARRA, S.A.  
Apartado, 1.217  
31080 PAMPLONA (Navarra)  
Tel. (948) 33 08 12  
Telex - 37786 E (IGNK)

A menudo, los investigadores tienen que tomar una decisión sobre la precisión con que la media de una muestra representa a la media de la población. Los intervalos de confianza -CI- pueden utilizarse para indicar la seguridad que uno tiene sobre una estimación dada. Por ejemplo, todos los datos entre un SEM de -1,96 y otro de 1,96 cubren el 95% de la muestra y representan entonces un CI del 95%. La probabilidad de seleccionar al azar un valor dentro de este intervalo es del 95% ( $P = 0,96$ ) y la de que se halle fuera de él del 5% ( $P = 0,05$ ). Y en el caso de movernos de un SEM de -2,58 hasta otro de 2,58 tendremos cubierto un CI del 99%.

En la investigación científica el empleo más corriente de las estadísticas es para comparar dos o más muestras para ver las diferencias entre ellas. Tales diferencias, de existir, no deberían provenir de unos errores de muestreo sino de los efectos de los tratamientos aplicados. O también puede llegarse a la hipótesis nula de que, no existiendo tales diferencias, los tratamientos experimentales no han diferido entre sí.

El valor utilizado cuando se comparan dos series de datos es el error estándar de la diferencia -SED- entre las medias. El SED es una distribución de diferencias entre las medias de muestras pareadas sobre una media de cero. Si la probabilidad es del 5% o menor ( $P \leq 0,05$ ), con la variación observada entre las medias -SED- ocurriendo al azar, esas medias serán significativamente diferentes. Si esta probabilidad es del 1% o menor ( $P \leq 0,01$ ), las diferencias entre las medias serán altamente significativas.

Para determinar si existen diferencias significativas entre las medias de unas muestras se utilizan o bien las estadísticas "z" -para grandes muestras normalmente distribuidas- o bien las "t" -para muestras pequeñas-. Entonces, basándose en las decisiones referentes a la hipótesis nula, pueden presentarse dos clases de errores, del Tipo I y del Tipo II. Un error del tipo I es el que tiene lugar cuando se halla una diferencia estadística entre las medias de las muestras -por ejemplo, cuando la hipótesis nula es rechazada- pero cuando tal diferencia se debe a un error de muestreo. En cambio, un error del Tipo II es aquél que se presenta cuando no se halla ninguna diferencia estadística entre las

medias de las muestras -cuando la hipótesis nula no se rechaza-, incluso cuando no existe diferencia entre las medias.

La hipótesis nula discutida en este punto se refiere sólo al caso de existir una diferencia entre dos series de tratamientos y no al caso de que un tratamiento sea mejor que otro. Lo primero es lo denominado hipótesis bifurcada ya que no se indica la dirección de las diferencias esperadas. Sin embargo, si una hipótesis nula indica que un tratamiento es mejor que otro, se utiliza una hipótesis unidireccional, la cual no requiere una diferencia tan grande entre las medias como la que requiere la anterior. Esto se debe a que los puntos críticos para una prueba bifurcada, siendo  $P = 0,05$ , son de una SD de 1,96, en tanto que para una prueba unidireccional son de una SD de 1,64.

Si se hace una comparación entre la variabilidad de una serie de datos contra otra se emplea la estadística F. En la prueba F la varianza es una medida de variabilidad más que la SD ya que las varianzas son aditivas y las SD no. Entonces se calcula una relación F para las dos series de datos a comparar, cotejándose con las tablas F para determinar si se rechaza o no la hipótesis nula.

Un empleo muy importante de la estadística de la prueba F en la investigación biológica es el análisis de la varianza -ANOVA. Esta constituye una técnica extremadamente útil ya que reparte la suma total de cuadrados de los desvíos, utilizados para calcular las SD, en sus partes componentes. Estas partes provienen o bien de las variables independientes en la experiencia -tratamientos, bloques, interacciones, etc.- o bien de errores de muestreo. El número de variables independientes u orígenes de variación que están incluidas en la ANOVA depende del diseño experimental. Cuando se compara la variabilidad entre las medias de 3 o más muestras, se calcula el cuadro de la media -MS- de entre los grupos y dentro de ellos. Este MS se calcula dividiendo la suma de cuadrados de aquél término por los grados de libertad. La relación entre los valores MS, el valor F observado, determina si la variabilidad en las puntuaciones medias de las muestras es demasiado grande para provenir de un error de muestreo y determina la suerte de la hipótesis nula.

Con ANOVA es posible determinar la exis-